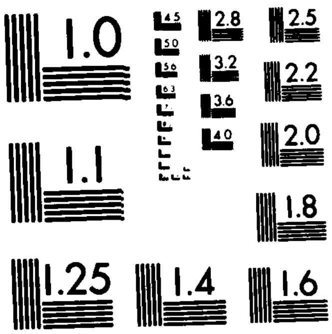
AD-A145 568 ANALYSIS OF DATA ENTRY PERFORMANCE: CHROMATIC VERSUS MONOCHROMATIC(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH R L ROSE MAY 84 1/3

UNCLASSIFIED AFIT/CI/NR-84-48D F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFIT/CI/NR 84-48D	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Analysis of Data Entry Performance: Chromatic versus Monochromatic	5. TYPE OF REPORT & PERIOD COVERED: ~~THESIS~~/DISSERTATION
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Reynold Lee Rose	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: AFIT STUDENT AT: Arizona State University	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: AFIT/NR WPAFB OH 45433	12. REPORT DATE: May 1984
	13. NUMBER OF PAGES: 221
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASS
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DTIC ELECTE SEP 17 1984 S D B

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

APPROVED FOR PUBLIC RELEASE: IAW AFR 190-1

Lynn Wolaver

LYNN E. WOLAVER
Dean for Research and Professional Development
AFIT, Wright-Patterson AFB OH

5 Sept 84

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ATTACHED

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASS

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD-A145 568

DTIC FILE COPY

84 09 13 016

ANALYSIS OF DATA ENTRY PERFORMANCE:

CHROMATIC VERSUS MONOCHROMATIC

by

Reynold Lee Rose

has been approved

March 1984

APPROVED:

_______________________, Chairperson

____

____

____

____

Supervisory Committee

ACCEPTED:

____
Department Chairperson

____
Dean, Graduate College

AFIT/CI/NR 84-48D

AFIT RESEARCH ASSESSMENT

The purpose of this questionnaire is to ascertain the value and/or contribution of research accomplished by students or faculty of the Air Force Institute of Technology (AU). It would be greatly appreciated if you would complete the following questionnaire and return it to:

AFIT/NR
Wright-Patterson AFB OH 45433

RESEARCH TITLE: Analysis of Data Entry Performance: Chromatic versus Monochromatic

AUTHOR: Reynold Lee Rose

RESEARCH ASSESSMENT QUESTIONS:

1. Did this research contribute to a current Air Force project?

( ) a. YES ( ) b. NO

2. Do you believe this research topic is significant enough that it would have been researched (or contracted) by your organization or another agency if AFIT had not?

( ) a. YES ( ) b. NO

3. The benefits of AFIT research can often be expressed by the equivalent value that your agency achieved/received by virtue of AFIT performing the research. Can you estimate what this research would have cost if it had been accomplished under contract or if it had been done in-house in terms of manpower and/or dollars?

( ) a. MAN-YEARS ________ ( ) b. $________

4. Often it is not possible to attach equivalent dollar values to research, although the results of the research may, in fact, be important. Whether or not you were able to establish an equivalent value for this research (3. above), what is your estimate of its significance?

( ) a. HIGHLY SIGNIFICANT ( ) b. SIGNIFICANT ( ) c. SLIGHTLY SIGNIFICANT ( ) d. OF NO SIGNIFICANCE

5. AFIT welcomes any further comments you may have on the above questions, or any additional details concerning the current application, future potential, or other value of this research. Please use the bottom part of this questionnaire for your statement(s).

NAME GRADE POSITION

ORGANIZATION LOCATION

STATEMENT(s):

ANALYSIS OF DATA ENTRY PERFORMANCE:

CHROMATIC VERSUS MONOCHROMATIC

by

Reynold Lee Rose

A Dissertation Presented In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY

May 1984

84 09 13 016

ABSTRACT

The effects of color versus monochrome video display terminals on the quality and speed performance of experienced data entry personnel were investigated. The ~~research~~ study considered operators experienced with the tested data entry task performed in a typical workforce environment. The task involved ~~the~~ computer entry of data for applicants requesting admission to Arizona State University. Nine operators with a minimum of 1.5 years experience on this job participated. All operators were female, ranging in age from 21 to 56 years. The dependent variables included objective measures of operator performance, speed and error rate, and a set of subjective measures of operator attitude. Four independent variables were considered: type of terminals (color and monochrome display), age of operator (35 years or less and greater than 35 years), experience level of operator (2 years or less and more than 2 years), and time of day of data entry (prior to noon and at or after noon). The research data were collected over a period of seventeen weeks. During this time 6688 items of data were collected.

The data were rigorously analyzed using the appropriate statistical methods. These methods included correlation, regression, and hypotheses testing through evaluation of the applicable statistics: ANOVA F, Welch W, or Wilcoxin T. The models investigated were 2-factor ANOVA models consisting of the terminal type variable in combination with each of the other independent variables: age, experience level, and time of day. These models were analyzed with respect to speed and error rate separately.

In brief, the results were that the attribute of color in the visual display used to accomplish data entry does not affect speed or error rate of the experienced operator. Subjectively, the attitude of the experienced operator was more supportive of ~~the~~ monochrome display than the color display.

To

Mary, Krysten, Mom, Dad & Jim

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

I. THE PROBLEM AND ITS BACKGROUND

Introduction

The information available to an organization is more than ever before being required in a timely fashion. To meet these needs, extensive computerized information systems are often seen as a requirement. Efficient use of such systems includes consideration of many aspects. One of the most important to the human factors engineer is the interface device. The popular interface devices are keyboard with hardcopy and keyboard with cathode ray tube (CRT) or visual display. Use of the visual display terminal as an interface tool is spreading rapidly and is becoming the primary communication tool between user and computer. When purchasing a video display terminal, the user has a basic issue to resolve: whether to acquire monochrome or color displays.

Background

Monochrome displays have been utilized since the 1960s and are technologically reliable. Color, however, is rather new on the scene. Color displays have been available since the early 1970s but did not take hold until the late 1970s. As a new technology, color displays are more costly than monochrome displays of similar

resolution and quality. Some of the computerized information system users choose monochrome to save these additional costs. Others make this choice because they feel monochrome can satisfy their needs quite adequately. One of the earliest manufacturers of color CRT display terminals contends that use of the color visual display enables the user "to organize data more logically and this leads to quicker and more accurate comprehension" (Myers, July, 1981, p.82). The company also states that they will only produce color display terminals as "color dramatically reduces operator fatigue and can cut costly errors by as much as 80%" (Myers, July, 1981, p.82).

The color terminal is gaining wide popularity in business and industry. It is being used extensively in hospitals, banking, design work, merchandising, data entry operations, and customer service operations. However empirical evidence and guidelines on its effective use are sparse in the literature with the reports very narrow in scope.

The literature can be categorized under one of three groups: unsupported comments, empirical research and user evaluations. The majority of the unsupported comments and all the user evaluations cited in the literature agreed

that use of a color display allows for more efficient processing of information and hence improves human productivity. The conclusions of the majority of the empirical research were based on analyses of performance of simple discrete tasks such as search, identify, locate, and count. The experiments were generally conducted in an isolated laboratory setting using subjects who were lacking in experience with the tested task (Christ, 1975). The alternatives to which color was compared were letter or numeral codes and geometric forms. Without exception the measures of performance were speed and/or accuracy. The findings in these studies were inconsistent with the unsupported comments and user evaluations. The empirical results can be summarized by stating that the relative effectiveness of color is a function of the task undertaken. Color generally was not found to have a unique advantage over coding information.

Christ (1975) suggested future research consider experienced operators, complex task and typical workforce environments. The study reported here was designed to include a complex task accomplished in a workforce environment using experienced operators in a data entry department. The desire was to conduct a controlled experiment to discover when, if ever, color terminals will

increase the effectiveness of data entry activities. Further, the desire was to determine how much improvement could be expected as a trade-off for the higher cost of color terminals. Such results are necessary if rational decisions are to be made relative to the acquisition of color terminals.

Problem Statement

The research reported here was designed to help fill the void of quantitative information on color display computer terminal use. Specifically, the desire was to see if the data entry time and error rate of experienced personnel would be reduced by converting to color terminals. The task was accomplished in a typical workforce environment.

Dissertation Outline

The remainder of this chapter discusses the chapters to follow.

Literature Search

A detailed description of all related literature is presented in Chapter II. These items are grouped as: unsupported comments, empirical studies, and user evaluations.

Methodology

In order to gain insight into the research problem,

a controlled experiment was designed to utilize a typical workforce environment in which experienced operators accomplished data entry. Chapter III details all aspects of the experiment in design and procedures. Initially the research problem was broken into subproblems to permit easier objective analysis from which conclusions and inferences could be made. An experiment was designed that incorporated these research questions with the characteristics of the data entry environment being tested. The experimental design is presented to include: the definition and justification of the variables, the data entry task studied, the subject population participating in the study, and equipment/support required. The experimental procedures followed are also discussed in detail.

Data Analysis and Results

The analysis of the data collected during this research study and the results are presented in five consecutive chapters.

Because of the complexity and extent of the data collected, Chapter IV discusses the process by which data were collected. The preparation required to format these data for analysis is also described.

Chapter V described the analyses accomplished on

data gathered when all operators were working on the monochrome display terminals. This portion of data assisted in setting the baseline for each operator used in the study and the analysis approach to pursue for the remaining data gathered.

Chapter VI describes the major analysis effort in this research. This chapter presents the analysis technique and approach to be applied and discusses each of the dependent measures of operator performance: accuracy and speed. The independent variables considered were terminal type, age of the operator, experience level of the operator, and time of day of data entry.

Chapter VII discusses the analysis and results of three special experiments conducted during the final three weeks of the study. These experiments were conducted to validate the effects of time on the results and to allow comparison of the current research to an earlier unpublished study.

Chapter VIII describes the analysis and results of the two survey instruments administered as a part of the research study. These were the single terminal evaluation survey and the multiple terminal comparison survey.

Conclusions and Recommendations

Chapter IX presents a summary of the overall

research results and conclusion. These conclusions are then generalized to the problem statement. The chapter and the research concludes with recommended areas for further study. These recommendations are based on the recognized limitations within the research, and the suggestions and voids in the literature.

II. LITERATURE SEARCH

Introduction

As the industry stands today, one needs not question whether color should be added to CRT displays. The technology is here with the hardware and software designers integrating this attribute into computer systems. The question which needs to be addressed is what are the advantages and drawbacks of using the color capability of a computer system (Hanson, 1979)? In particular, does the use of color display terminals enhance the productivity of experienced data entry operators whose productivity is measured in terms of speed and error rate? The literature was searched in attempts to gain insight into this question. This search included all that had been reported on the use of color in visual displays without regards to the kind of display interface, the type of task or the experience of the operator. The literature was categorized into three groups. Some authors make general comments about color and the use of visual displays but offer no empirical evidence for their conjectures. Other researchers review and comment on studies they themselves or others have done as well as present findings from their own empirical research. Finally current users comment on the impact color

terminals have had on the productivity in their organizations. This literature review chapter is developed according to these three categories. Initially the findings in each category are reported and summarized. Then an overall summary is offered. Based on this summary, the voids in the knowledge on how color display terminals enhance productivity are highlighted.

Literature Search

Introduction

The literature search utilized the Arizona State University (ASU) computerized literature search service. The data bases queried included INSPEC, NTIS, COMPENDEX, ERIC, Dissertation Abstracts and Microcomputer Index. Abstracts from the queries were reviewed to determine the source documents appropriateness. Potentially useful papers were located and read. Relevant references from these reports were also examined. Where possible the source document for each study found particularly key to this research was found and critically reviewed. In addition personal communications were established between the researcher and several noted authors in the Human Factors field of color research: Richard Christ, Thomas Tullis and H. Rudy Ramsey. Their personal critically annotated bibliographies were provided to complete the

literature search. Personal communication was also established between the organizations which published using color terminals in their organizations for data entry. Three of the companies responded by making available their unpublished detailed reports.

This search resulted in reports on what others feel or have investigated on the effectiveness of color in visual displays. These are detailed next according to the three categories: unsupported comments, empirical research and user evaluations. Those studies of particular interest to this research are expounded upon in more detail than the others.

Unsupported Comments

Conjectures. The literature reports a number of positive though empirically unsubstantiated justifications which users are giving for choosing color CRT terminals. Color is more aesthetically pleasing (Calev, 1981; Murlijacic, 1981; Myers, July 1981). Color holds a natural attraction for human beings as they observe the majority of their surroundings in color (Horsley, 1981; Myers, July 1981; Taylor, 1979). Color is easier to read (Information handling: The major trends, 1980), more interesting, and more informative (Buchanan, 1979; Horsley, 1981). Color can draw attention, emphasize, and

stimulate (Belie and Rapagnani, 1981; Durrett and Trezona, 1982; Myers, July 1981). Color speeds identification, improves visibility, and reduces response time (Friend, 1980; Morris, 1979; Whieldon, 1981). All of these comments are presented without empirical evidence provided.

Some of the authors who write about the use of CRT displays in computerized information systems offer comments against the use of a color CRT. They favor the use of a monochrome CRT. Monochrome can do many jobs quite adequately while color can be overused or used incorrectly (Myers, June 1981; Truckenbroad, 1981). Color used in large display densities, more than thirty items, and in large color codes, more than six colors, can increase search times (Carter and Cahill, 1979). More than eight colors slow operator response time (Morris, 1979). Color is second only to adequate contrast and display clarity, as a matter of personal preference (Cakir, Hart, and Stewart, 1980) and can be a hindrance to performance (Bylander, 1979). All of these indicate that monochrome can sometimes prove as adequate or even superior to color.

<u>Summary</u>. Unsupported comments on the effects color in visual displays have on the user are numerous in the literature. The majority favor color, purporting it

improves productivity. Others are more conservative, contending that a monochrome display is as adequate or even superior.

Empirical Research

Research in the area of color coding has been a topic of interest since the 1950's. Most of this research has been concerned with identification and search tasks and have used random patterns of letters , numbers, and geometric shapes. The literature hosts a number of researchers who reviewed and commented on studies accomplished by themselves or others as well as present findings from their own experiments. The reviews promote the idea that color, when used properly and for certain tasks, can be superior to monochrome. These findings are reported next in the chronological order in which they were done, beginning with the earliest. Those which have particular relevance to this research are reported and criticized in greater detail than the others.

Objective Literature. Green and Anderson (1956) performed research using operators whose work involved the use of control panels. The purpose of the study was to investigate the effectiveness of color coding as a function of the relationship of the number of symbols of each color and the number of colors used. Kodachrome transparencies were used for the study and were projected

onto a screen ten feet from the operator. The colors used were green, red, yellow, and blue. A flashlight pointer was used to have the operator indicate the desired target. Search time and error rate were used as measures of performance. It was concluded that when the operator knew the color of the target, search time was proportional to the number of symbols of that color. However, if the color of the target was unknown, the search time was found to be proportional to the total number of symbols displayed.

Christner and Ray (1961) accomplished a research study to determine the relative effectiveness of selected target background coding combinations using a map reading task. Three target codes (color, number, and enclosed shape) and five backgrounds (all white, solid gray, five shades of gray, five pastel hues, and five different patterns) were used. Identifying, locating, and counting tasks were involved. Acetate overlays were used to display the maps on a 30x30 inch piece of cardboard shaded according to one of the five backgrounds specified. Five operators were involved in the experiment. For the identifying task, number coding was found to be superior to color coding. For both the locate and count tasks, color coding was superior to number coding. This was one

of the first studies that clearly illustrated the task dependency of improved performance due to color coding.

In 1961, Hitt studied the relative effectiveness of selected abstract coding methods based upon effects on various operator tasks. The measure of operator performance used was the number of correct responses per experimental condition and the speed of response. The five coding methods were numeral, letter, geometric shape, configuration, and color. The colors used were black, red, blue, brown, yellow, green, purple, and orange. Identify, locate, count, compare, and verify tasks were studied. The subjects were tested for both visual acuity and color blindness prior to the first treatment. The treatment was a series of 30x20 inch cardboard posters divided into forty cells each. The targets were randomly placed in the forty cells at varying levels of density. Pencil and paper tests were administered to the two groups of five subjects to measure the five tasks. Color and numeric coding were found to be the two superior coding methods. There was no significant difference found between color and numeric coding except in the identify task, where numeric coding was found to be superior to color coding.

Schutz (1961) accomplished research with the primary

purpose to determine the effect of multiple line versus multiple graph presentation of trend type graphical displays on operator performance. Color and black/white coding were both used in this study. The colors used were red, yellow, green, and purple. In the black/white presentations, four different line codes were presented. The graphs were displayed via projection of 35mm slides on a screen. The capability existed to project up to four 5x5 inch graphs on the screen at one time. Both a point reading and a comparison task were utilized. The point reading task consisted of reading from a given line graph a vertical axis value associated with a specified horizontal axis value. The comparison task involved determining from multiple graphs or graphs with multiple lines the highest vertical axis value associated with a particular horizontal axis value. Ten male subjects were utilized in this study. In general it was concluded that color improves performance for the point reading task but not for the comparison task. Once again the use of color to improve performance was shown to be task dependent.

Jones (1962) commented on eight studies accomplished using slides or hardcopy stimulus. He states that "color codes do not appear to be suited for situations that demand rapid and precise identification, whereas they are

valuable in decreasing search times with the locate-type tasks" (p. 355).

S. L. Smith performed a series of three studies in the 1960's involving primarily a search task. Each of these studies measured search and counting times and counting errors with respect to various color presentations of different classes of targets. These presentations were made using a series of 2x2 inch slides and a rear projection system. The first study in 1962 involved twelve subjects and a visual search of 300 displays. The projection display was twelve inches square and viewed by the subjects at a distance of eighteen inches. The colors used were red, green, blue, orange, and white. Searches were made by the subjects both knowing and not knowing the color to be used for the target. Neither the particular color of the target nor the display background had any significant effect on search time. For multicolored displays, if the color of the target was known in advance by the subject, search times were considerably shorter than when the target color was unknown.

In 1963, Smith conducted another study to further consider the effects of color as a redundant code. This study used a 32 inch square display field presented to the

subjects at a distance of four to five feet. The colors considered were green, white, blue, red, and yellow. As before the color of the shapes presented or the background did not cause a significant variation in the measures of search and counting times or counting errors. However, when color was used as a redundant code, a 65% reduction in search time and a 69% reduction in counting time were recorded. Along with these was a 76% reduction in counting errors.

A third study was conducted by Smith and published with the assistance of Thomas in 1964. This study was an "attempt to measure systematically the superiority of display color coding, by comparing it with various shape codes" (pg. 138). The colors of interest were green, blue, white, red, and yellow. The shape codes were military symbols, geometric forms, and aircraft shapes. Eight men and women with normal color vision each reviewed 550 displays over four experimental sessions. Subjects were asked to count the number of occurrences of a particular target class from a 29 inch square display viewed at a distance of five feet. Count time and number of errors were measured. It was found that colors were counted about twice as fast for the easiest set of symbols and about three times as fast for the hardest set of

symbols. Also supported was that if in color the count times for each of the shapes were not significantly different. Fewer errors were made in color counting than in shape counting. This series of experiments presents some of the areas in which color usage has been found to have an effect on subject performance.

Brooks (1965) also studied the effect of color coding on search times. Six groups of ten subjects each were asked to respond to ten different displays containing 60 symbols, some of which were color coded. The displays were on standard size bond paper mounted to a clipboard. These displays were presented to the subjects at a distance of approximately eighteen inches. The symbols were the letters H, S, F, I, and M followed by a three digit number. The colors used were red, yellow, green, blue, and violet. When each display was presented, the subject recorded as quickly as possible the ten numbers associated with the letter H. If color was used, the subjects were told that the letter H would have a red rectangular bar beneath it. The displays viewed by group one had no color. Each successive group had the addition of a color bar in one of the colors under all occurrences of one of the letters. For group two, only the red bar was presented and it was under the H. The color bar added

for group three was yellow below the F; for group four it was green below the I; for group five it was blue below the S; and for group six it was violet below the M. Color was found to be significantly better than the no color condition. The increasing number of colors from group two to group six had no significant affect on the subject's search time provided the subject was told to search for an item presented in a particular color. These findings are the same as those from the study accomplished by Smith and Thomas (1964).

Smith, Farquhar, and Thomas (1965) designed an experiment to assess and compare the effects of symbolic, numeric and color coding in formatted displays. The displays consisted of two digit items presented in tabular matrix format. Each matrix had ten rows and either two, six, or ten columns. A rear projection system was used and the displays appeared as white or colored figures on a black background. The 25 inch square display was viewed by the subjects at a distance of five feet. Twelve subjects were tested using the tasks of row comparison and item counting. The colors used were white, yellow, red, blue, and green. For the row comparison task, relevant color coding significantly improved the subjects' performance time and decreased their error rate over any

other of the code conditions. Frequently in the item counting task the counting time and error rate of the subjects were significantly improved for displays with relevant color coding.

In 1968, Munns studied the effect of varying certain aspects of displayed symbols upon operator performance. The display used consisted of a series of 8x10 inch problem sheets inserted into a 45 degree sloping panel to simulate a military radar display tube. Twelve male subjects were requested to use the simulated radar console to detect enemy aircraft and assign interceptors. Blue and red colors were added to the displays as a redundant code. Although color reduced performance time, there was no indication that color would reduce error rate. The subjects reported "feeling more secure with color" (pg. 1221). Some other comments made by the subjects were "color makes it easier" and "color helps" (pg.1221).

Chase (1970) accomplished a study to determine the effects of several variations of two types of visual display systems on subjective pilot evaluation and objective measures of performance in landing approaches. The landing approaches were simulated with either a projector or a collimated TV display system. The variables of interest were color, differences between

displays due to collimation, and reduced resolution. Seven professional pilots participated in the study. The pilots were critical of the black and white variation of either display and favored use of a color system. The advantages cited for a color system included greater pilot relaxation, decreased fatigue, better picture quality, and more realistic depth perception. It was also concluded that it took more concentration and effort to fly without color. More eyestrain and blinking were noticeable with the black and white displays. Also from the performance measures of speed and error rate, visual cues were perceived more quickly when using the color displays.

Dooley and Harkins (1970) experimented with learning and attention effects of color when used as an information code or decoration on column charts. The charts used were presented in hardcopy form. The colors investigated were red, green, and blue. Three groups of fifteen subjects participated in the study. Although it is generally assumed that the use of color in visual communications has an over all positive effect on learning and performance, it was found that black and white code was equally effective as an information transmitter. Color's principal effect was found to be in the area of motivational qualities.

Kanarick and Petersen (1971) conducted a study to determine whether keeping-track performance, especially of low valued items, could be enhanced by redundantly color coding the information. Fifty six subjects viewed a row of ten solid state readouts in which a number from one through six in one of six colors were displayed. The colors used were blue, violet, yellow, red, orange, and green. The subjects were required to remember the numbers as they were presented. The redundant color coding did not facilitate the keeping-track performance.

An experiment associated with a map reading task was accomplished by Shontz, Trumm, and Williams (1971). Thirty three junior and senior students enrolled in pilot and navigator programs of the Air Force Reserve Officer Training Corps at the University of Minnesota were asked to view achromatic or chromatic maps. The viewing was accomplished with the subject sitting in a chair looking up at a map that was concealed behind mechanical doors operated by the subject. The general finding was that color coding for information location is effective. The degree of effectiveness was found to be dependent on the number of categories coded, discrimination in the peripheral vision of colors used, and the number of objects per code category.

Wedell and Alden (1973) studied the effectiveness of color code versus numeric code on the keeping-track task of the air traffic controller. It was hypothesized that color was superior to numeric coding particularly with a greater number of total items displayed. A display system consisting of 35mm color slides rear projected onto a ground glass screen was used. This produced a 10 inch square display which was viewed by the subjects at a distance of 22 inches. Thirty six male subjects participated in the study and were pretested for normal color vision using the Farnsworth D-15 test. Different levels of aircraft load and numeric and color coding conditions were investigated. The colors used were red, blue, orange, green, yellow, and purple. The task involved detecting changes from slide to slide. Color coding was not found to be superior to numeric coding in either higher aircraft load or higher interrogation load. Color was useful for retaining the number of aircraft at a specific altitude. The results of error analysis indicated that color can aid in retaining information concerning the number of items presented, but identification information was quickly lost.

Cahill and Carter (1976) accomplished a study concerned with search task performance based on display

density and number of colors used in coding the display. Three digit numbers were displayed using a rear projection system with density between ten and fifty items per display. From one to ten colors were used to code the items. Twenty subjects participated in the study. An initial drop in search times was observed as the first few colors were added to an uncoded display. This was followed by a rise in search times as more colors were added to the display.

A study accomplished by Christ (1975) had as a purpose to determine whether the cost of converting military aircraft controller displays to color could be justified in terms of increased human information processing performance. The research began in 1973 under a contract with the Joint Army Navy Aircraft Instrumentation Research Working Group. The first step in Christ's research included an extensive in-depth review of the literature to determine the knowledge that was available which was relevant to the project. Based on the 42 studies he found which satisfied his objectives, he reported glaring gaps in the quantitative information of the affects of color in visual displays. The key criticism concerned the subjects used in the studies to date. They were all inexperienced with the task under

study. Christ also pointed out that all the research were done in isolated laboratory settings where distractions were negligible. The task accomplished was a relatively simple discrete task and the subjects were allowed to focus total attention to accomplishing it.

Based on the purpose of their research and their discovered voids of knowledge in the literature, Christ and his associates, Teichner and Corso, designed a series of nine experiments. They included two ranges of experienced operators, isolated as well as integrated laboratory settings and a variety of tasks which varied in complexity.

Eight males who had practiced the task to proficiency over nine months as well as eight others who had trained for a shorter time were used as subjects. The environments included the subjects accomplishing a simple discrete task in isolation, a series of simple tasks given in random order and a series of simple tasks given in random order while the subject assumed the duties of an air traffic controller. Letters, digits, familiar geometric shapes and colored dots were employed as the coding dimensions in the visual displays. The codes were: letters (C, H, K, N, P, S), digits (2, 3, 4, 5, 6, 7), geometric shapes (circle, square, triangle, diamond,

cross, star) and colored dots (purple, green, yellow, orange, red). The treatments were projected onto a display screen with the colored dots created by using colored filters. The tasks included choice reaction time which required no memory, search and locate, information-memory, and a same-difference task. Response time and accuracy were the measures of performance.

After the subjects were trained to proficiency with the task, there were no clear and consistent advantages for any one of the visual code sets over the others. It was concluded that the relative effectiveness of different visual codes varies as a function of practice. If long term performance increase is an objective, the code symbology used in a visual display is irrelevant. If an increase in short term performance is the object and if relatively inexperienced operators are to be utilized, manipulation of visual code sets may improve performance.

The overall conclusions from all of their research was concurrent with their literature findings. The use of color as an information code had no different affects on user performance than the use of monochrome. If there was a difference, it was minimal and tended to disappear with practice. Color was sometimes associated with the best performance and other times with the worst performance

(Christ and Corso, 1983; Teichner, Christ and Corso, 1977)

Ohlsson, Nilsson, and Ronnberg (1981) studied speed and accuracy in a scanning task as a function of combinations of various text and background colors. The text colors considered were cyan, green, yellow, magenta, red, white, and blue along with the various background colors of cyan, green, yellow, magenta, red, white, blue, and black. Eight subjects ranging in age from 20 to 30 years participated in the study. The display used consisted of a TV monitor on which letters were presented in two matrices of size nine rows by nineteen columns. The subjects were asked to scan these rows as quickly as possible and count the number of occurrences of a particular target letter. The findings, in terms of average speed and accuracy were then presented for each of the text/background combinations considered.

Tullis (1981) studied narrative, structured, black and white graphics and color graphics treatments in a computer aided decision making task. The research was accomplished using eight Bell System employees. Each employee was subjected to the various treatments over a period of seven hours. The five male and three female subjects ranged in age from 25 to 50 years and had from .75 to 3 years of computer experience. It was found that

the two graphics formats were superior to the narrative formats. There was a lack of significant difference between the color and the black and white graphics formats. A questionnaire was administered in which seven of the eight subjects chose color graphics over black and white graphics as the one they would prefer to work with on a daily basis.

Keister (1981) was one of the first found who investigated other than the typical search, locate and count type tasks. He considered color applied to data entry tasks to determine if "throughput and color accuracy were effected" (pg. 736). This study involved two experiments using an INTECOLOR 8051 computer with eight colors. The first experiment involved eight females experienced with typing and data entry but not familiar with the task to be accomplished in the experiment. Each operator performed 80 entry or change operations of products containing five data fields on a display that was formatted to handle input of up to ten products before clearing the screen. Twenty of the entries and twenty of the changes were performed by each operator using the color display. In this display color was used for additional emphasis and to code the types of data. The other forty entries and changes were accomplished by each

operator using a monochrome display. This was a green phosphor display with reverse video for emphasis. Time per transaction and errors were the measures used in the experiment. None of the differences found in using the color display versus the monochrome display were significant. However, there were indications "that color speeds the initial learning of new entry tasks and that the effects of color are stronger in more complex tasks" (pg. 737). The second experiment involved 54 volunteers from National Cash Register (NCR) software development groups accomplishing only the change task for the monochrome and color conditions. Once again these subjects were unfamiliar with the task they were to perform. The outcome was similar to experiment one in that the overall effect of color was not significant. "However, there was a clear trend toward the superiority of color" (pg. 738). The following three comments were the general conclusions from this set of experiments:

> (1) In simple entry tasks, which tend to be rather boring, color can increase operator motivation and facilitate performance by making the task more interesting.
>
> (2) In moderately complex entry tasks, there is enough challenge to prevent serious problems with boredom. At the same time, the tasks are not so difficult that color coding is of particular value for aiding in locating items on the screen and keeping track of entry activities. Thus, for such tasks, color

> displays provide only a minimal advantage.
>
> (3) In more complex tasks, boredom is rarely a factor, but there is a tendency for operators to become confused. Color results in improved performance by making it easier for Ss (subjects) to keep track of their activities and to locate items on the screen (Keister, 1981, p. 739).

Another experiment concerned with data entry is an unpublished study accomplished in 1982 for ITT Courier by John D. Shafer, Productivity Consultant. This study utilized the ITT Courier 2790-2A color display terminal and was accomplished at Pennsylvania Blue Shield, Medicare Claims Division, Camp Hill, Pennsylvania. The application consisted of processing Medicare claims for payment. The study involved a subjective survey questionnaire as well as timing of task completion under controlled observation. The subjective survey involved twenty three operators who used the color terminals for a period of two weeks. These subjects had been using monochrome terminals for approximately one and one half years. For the objective part of the study, ten subjects were randomly selected from the twenty three who participated in the subjective survey. These ten operators had at least ten months experience with an average of sixteen months in the application used in the study. For training purposes, ten terminals were installed at Pennsylvania Blue Shield six

working days prior to data collection. Time for processing a claim was collected on each of the subjects for a period of three working days using both the color and monochrome modes of the new terminals. In the color mode, the four colors of blue, red, green, and white were used to code the various input and error fields. The "monochrome" mode was accomplished by turning the color switch off on the terminal which left the basic green color but also retained the white for error messages. It was felt by the author of this study that the color versus monochrome differences might be understated due to this retention of the second color. On the first day of the study, the operators worked half a day using the color mode and half a day using the "monochrome" mode. The second day was all in the color mode and the third day all in the "monochrome" mode. Timing of the inputs was accomplished by having each operator work their standard eight hour day and having an observer log any work stoppages. These stoppages were subtracted from the operator work time to arrive at a net processing time. During the three days a total of 10,645 claims were processed over a net total of 183.9 hours. The overall finding was that the color mode increased productivity by approximately eight percent.

Summary of the Experimental Literature. A majority of the studies from the literature involved identify, locate, count and search type tasks in a variety of applications. These tasks were explored while using a number of different coding methods: color, letters, numerals, geometric shapes, and others. A number of different display conditions were investigated such as density, formatted displays, and foreground versus background colors. Another area of study mentioned several times involved graphical presentation of information in color or monochrome. Both multiple graphs and multiple lines have been considered by researchers. A wide range of display methods were used in these studies. The general conclusion of the studies discussed was that color is effective in some situations, but detrimental in others. The research which reported on the data entry task differed slightly from these conclusions. Color either had minimal or a positive effect on data entry performance.

User Evaluations

Organizations. For the task of data entry several managers have commented in the literature about the use of color terminals in their organizations. A. Cessana and Associates are convinced of the value of color over

monochrome terminals. The marketing department manager states that their operators are at least 25% more productive. Reading and identification are easier using color displays (Whieldon, 1981). A. Cessana and Associates were personally contacted via letter and phone requesting the study and data supporting their evaluations. They were unwilling to make such information available.

Another user of color display terminals is the Morristown Memorial Hospital in Morristown, New Jersey. They are using the IBM 3279 four color (red, green, blue, and white) alphanumeric terminals in their admissions department. Eight of these units were introduced as replacements to some existing monochrome units in the admissions department in October 1979. Currently there are 25 operators ranging in age from 20 to 74 years utilizing the terminals to control admissions, records, and billing of patients. The majority of these operators are female with 24 of them wearing glasses. None of the operators are color blind. The operators enjoy using the color terminals and feel that eyestrain has been reduced. The Director of Admissions and Director of Systems and Data Processing both agree that since installation of the color units productivity has increased (Driscoll, 1983;

Miller, 1982).

The Wilkens Pipe and Supply Company, Peoria, Illinois also utilize the IBM 3279 terminals. Eight of the firm's 47 on-line CRT terminals are color units. These units are used for order entry, stock status inquiry, billing, purchasing, and receiving. The people really enjoy the color according to the Data Processing Manager (Kelso, 1983; Miller, 1982). The terminals are used to display data in color and highlight errors. The result is fewer order entry errors, higher operator productivity, and less pressure for the terminal operators. The "operators also claim an easing of eye strain and a relief from the monotony of reviewing a screen full of monotone data" (Color CRT terminals reduce error rates, 1981, pg. 83). These users agree that the use of color alphanumeric terminals increase operator productivity but no statistical evidence is given to support these comments.

Summary. The user evaluations of the effects of using color CRT displays are consistent with the unsupported comments discussed earlier. They all claim productivity is increased as well as eyestrain is lessened.

Conclusions

There continues to exist a paucity of quantitative knowledge and empirical evidence in the literature on the effects of color on performance of other than search, locate, and count type tasks. Even for these tasks, research is scarce that incorporated personnel experienced with the task, tasks that are complex, and tasks that are done in environments which emulate those of the workforce. In particular, there is an absence of information on the effects on operators who are very experienced with the task of data entry accomplished in a realistic workforce environment. This information is essential to organizations whose goal is efficient use of resources. With this knowledge, human factors engineers can satisfy one of their key objectives of designing a human/machine interface which allows for maximum human performance.

The significant research question of current interest is the effect of the use of a color alphanumeric terminal versus a monochrome alphanumeric terminal on operator productivity when performing a familiar data input task in a workforce environment? When considering this question a number of other variables are also of concern. What is the effect of the use of color terminals on operator eyestrain and attitude toward the job? Does age,

experience level, or time of day of data entry have an effect on the above variables of interest? These questions are the primary concern of this research.

III. METHODOLOGY

Introduction

As cited in the literature search, Chapter II, Christ's (1975) three key suggestions to be considered in future research concerned with color visual displays were (1) use of operators experienced with the tested task, (2) the task should be complex and (3) the task should be accomplished in a workforce environment. With these suggestions in mind an experiment was designed. The research question was subdivided into subquestions to permit easier objective analyses from which conclusions and inferences could be made. The experiment incorporated these research questions within the characteristics of the data entry environment as described in the literature and detailed in Chapter II. This chapter details the research questions, the experiment, and concludes with a discussion of the procedures used to perform the experiment.

Research Questions

The significant research question under investigation was the effects of usage of a color display computer terminal versus a monochrome display computer terminal on operator productivity when performing a familiar data entry task in a workforce environment. More specifically,

does color display usage affect input error rate? Does color display usage affect the time required for data entry? Are the effects of color display usage related to the age and/or experience level of the operator? Are the effects of color display usage related to the time of day of data entry? Do these effects remain the same over an extended period of time? Do the operators feel that the computer terminal type, color or monochrome display, used for data entry influences job satisfaction and/or how well the operator likes the terminal? Does terminal type influence the effects that interruptions have on the operator when they are performing the data entry task? Are eyestrain and headaches a problem for the operator when working with either of the two terminal types? Is physical fatigue a problem for the operators when working on one type of terminal versus another?

In order to investigate these research questions in a meaningful way, the characteristics associated with the data entry workforce environment and previous experiments were identified via the literature. An experiment was designed that incorporated these characteristics and allowed insight into the stated research questions. The experiment is discussed next in detail.

Experimental Design

Introduction

An experiment that allowed objective consideration of the stated research questions was designed with characteristics emulating those associated with the data entry environment cited in the literature. This experiment is described by considering four factors: identification of the variables, the data entry task performed, the population sampled, and the equipment/support required. Discussion of each of these factors includes stating, where applicable, the characteristics of the data entry workforce environment reported in the literature. Included in these citings are two companies who use both color and monochrome display computer terminals for data entry. One is the Wilkens Pipe and Supply Company, a large wholesale plumbing supply firm based in Peoria, Illinois, subsequently referred to as User Company A (Color CRT terminals reduce error rate, 1981; Kelso, 1983; Miller, 1982). The second is the Morristown Memorial Hospital in New Jersey, subsequently referred to as User Company B (Driscoll, 1983; Miller, 1982). In addition to the two companies, characteristics of two previous research studies are detailed. One study was performed by Keister in 1981, herein referred to as

the Keister study. The other was an unpublished study accomplished for a company that manufactures IBM compatible interfaces by a productivity consultant (Shafer, 1982). This latter study is subsequently referenced as a consultant study. Additional literature is cited as required. The characteristics of the two user companies and the two previous studies are then related to those of the research reported here for each of the factors.

Variables

When designing the experiment three categories of variables were of interest: dependent or measured variables, independent or controlled variables, and exogenous or uncontrolled variables. The dependent variables were objective measures of operator performance (speed and error rate) and a set of subjective measures of operator attitude (job satisfaction, terminal preference, effects of interruptions, glare, eyestrain, headaches and physical fatigue). The independent variables included those which the literature, or personal observation, has suggested could affect data entry performance, either individually or in combinations. Those of concern to this study include terminal type, operator age, operator experience level, and time of day of data entry. The

exogenous variables are those which were measured but not controlled by the experimenter. These included room lighting, room temperature, room humidity, and operator color detection deficiency. Each variable is discussed from the standpoint of the literature and then operationally defined for the current research experiment.

Dependent Variables. The first dependent variables were objective measures of operator performance. The literature suggests that the objective measures of operator performance of a data entry task, accomplished via computer terminal, are time to perform the task and error rate (Color CRT terminals reduce error rates, 1981; Driscoll, 1983; Keister, 1981). User Company A studied the performance of their operators on time to complete a transaction and the number of errors made. User Company B similarly studied operators accomplishing data entry in their admissions office. Daily reports were written by the Director of Admissions of User Company B identifying operator performance. The measures used were also time per new patient data entry and number of errors made. The Keister study states that "time per transaction and errors were the basic measures" (p. 737) of operator performance. However his analysis investigated only the time variable. Time was defined as the number of seconds to complete the

data entry task successfully. Successfully meant with no errors remaining. Like Keister, a consultant study considered and defined time similarly. In addition, the task was not completed until it was done correctly. Implicitly this assumes time and errors are highly correlated, therefore when the time variable was analyzed, the conclusions were consistently generalized to errors.

The experiment reported herein also used time to complete the task and error rate as measures of operator performance in a data entry operation. The two variables, session time and error count, were defined for this research study. Session time was defined as the clock time in seconds required for an operator to enter one new application. A "new application" included any form prepared via computer terminal on a person applying for admission to Arizona State University (ASU) for which no information existed on the computer system at the time the session began. Error was defined, similarly to Altman's (1964) definition, as any operator act which the computer recognized as incorrect and therefore would not allow task completion. As in other studies, all errors had to be corrected before a session was completed. Error count was the number of errors made by the operator during a given session.

The other dependent variable considered was a set of subjective measures of operator attitude. Some subjective comments with respect to color are reported in the literature. The operators of User Company A claim an easing of eyestrain (with color) and a relief from the monotony of reviewing a screen full of monotone data. The operators of User Company B registered glare as a problem with both color and monochrome terminal displays. The operators in User Company B also reported that headaches and eyestrain were greater when using the monochrome displays. A consultant study produced comments on a subjective survey such as "easier to see errors", "less eyestrain", "less fatigue", "increase productivity", and "reduce errors".

The current study investigated operator attitude collectively using two survey instruments. The instruments measured operator reaction to factors such as job satisfaction, terminal preference, effects of interruptions, glare, eyestrain, headaches and physical fatigue. The instruments are discussed in detail later in this chapter under the heading Miscellaneous Measuring Devices/Equipment.

Independent Variables. Based on the literature and personal observation, several independent variables were

included in the experimental design. These included terminal type, operator age, operator experience level and time of day of data entry. Each of these are discussed separately relative to both the literature and the current research.

The literature reporting on the effects of color on operator performance always considered two types of computer terminal displays: color and monochrome. User Company A compared the IBM 3279 four-color terminal to an unidentified monochrome terminal. Similarly, User Company B also compared the IBM 3279 four-color terminal and a monochrome terminal. The Keister study employed the INTECOLOR 8051 computer terminal with eight colors for his comparison experiments. A consultant study used their make 2790 computer terminal with four colors. For the monochrome alternative, a consultant study used the same 2790 terminal but with the color switch in the off position.

The research reported herein considered two types of displays: monochrome and color, consistent with earlier research studies.

Previous data entry studies did not consider the independent variable of operator age. However, the literature review did uncover in several studies concerned

with age in which operator performance on selected tasks was significantly related to age. One study which investigated age as a factor in combined manual and decision tasks detected significant differences between the performance of the younger (18 to 29 years) and the older (52 to 63 years) group of operators (Kochhar, 1979). Another study found that after 20 years of age, attitude of workers dropped steadily until about 35 and then began to rise steadily (Kunze, 1975). Two levels of operator age were used for the current study. One level consisted of those operators 35 years of age or less and the other, those operators greater than 35 years. The operator ages were calculated as of 7 February 1983, the beginning of the experiment.

The variable of operator experience in data entry tasks was not considered in cited research studies. However, in an earlier study of clerical workers it was found that lack of experience appeared to be the chief cause of poor worker performance (Kunze, 1975). Christ (1975) has emphasized the glaring gap of rigorous research in the area of color when experienced operators are performing the task under study.

In the current research, operator experience was included as a variable to weight the number of years an

operator had been performing the tested data entry task via computer terminal. The Director of Undergraduate Admissions at Arizona State University (ASU) has theorized that it takes an operator a minimum of two years to become totally experienced with the prescribed data entry task on a computer terminal (Neary, 1983). The variable of operator experience was considered at two levels: less than 2 years experience and 2 years or more experience.

The time of day of data entry was also not considered by data entry studies in the literature. However, "there is considerable evidence that human performance varies during the course of the working day" (Craig, 1979, p. 61). Hence this variable was of interest to the current study. The time of day was defined as that time at which the operator began a data entry session. Since there is evidence to support there is a drop in performance following lunch (Blake, 1967), this time was used to split the variable into two levels. The levels were morning and afternoon. Morning was any session that began prior to twelve o'clock noon. Afternoon was any session that began at or after twelve o'clock noon.

Exogenous Variables. Several environmental and operator variables were measured in the study but not manipulated by the experiment. The environmental

variables of room lighting, temperature and humidity were measured periodically during the experiment to insure that these variables remained constant throughout the course of the experiment. This was deemed necessary because of the many weeks of time over which data were collected. Relative to the experimental subjects for the experiment, color detection deficiency was checked to insure that none were color blind. This was necessary since color was the independent variable of primary concern to the research. These measures are described in detail later in this section under Miscellaneous Measuring Devices/Equipment and in the Experimental Procedures section under Data Collection-Objective.

Data Entry Task

The data entry task used in this experiment was selected to emulate data entry task characteristics cited in the literature.

Literature Reported Task. The data entry task accomplished by User Company A was an ordering process. The orders for various plumbing supplies were received and the data entered into the computer via a terminal interface. This entry was accomplished on a screen presented order form. The order form filled the majority of the display. The task was accomplished daily by each

of the operators (Kelso, 1983).

The data entry task accomplished by User Company B was the entering of information on patients being admitted to a medical facility. This task also used a screen presented form. A full screen of information was required on each patient. This was a daily task accomplished by each of the operators.

The characteristics of the data entry task used in earlier research studies was also similar to those used by the User Companies. The Keister study used an ordering-form task similar to that of User Company A. A consultant study used an insurance claims task. This latter task required the operator to enter a full screen of information on a preprogrammed form presented via the computer terminal display.

<u>Current Experiment Task.</u> The data entry task used in the research reported herein was similar to the tasks just described and adaptable to the selected site, the Undergraduate Admissions Office at Arizona State University. The task involved the entry of data for "new applicants" to the University. A preprogrammed form, presented on a monochrome display computer terminal and used for the past 4 years for this task, was used for data entry. The form required input of a full screen of

information for each of the new applicants (Appendix A). The information entered was similar in quantity for each application. Each experimental subject performed the task on a daily basis.

Population

The population involved in the current research was selected to have characteristics of the general workforce performing data entry tasks. This was desired in order to allow generalization of results from the current research to users with similar characteristics. The characteristics of the population are discussed by first presenting those derived from the literature and then adapting them to the reported study.

User Company A employed operators whose experience with the data entry task ranged from 1 to 5 years and whose ages ranged from 19 to 62 years. Women made up 88% of the workforce accomplishing data entry. User Company B operators performing data entry similarly had experience with the task ranging from 1 to 5 years and ages between 20 and 74 years. Women comprised 80% of these operators. In the Keister study, eight operators were employed in the primary data entry experiment. In a consultant study, ten people were studied as subjects. Both of these later studies used all females operators.

The characteristics of the population in the reported study were similar. The experience of the operators ranged from 1.5 to 3.5 years. The ages of the operators was from 21 to 56 years. All of the nine operators involved in the current research were female, employed by the Undergraduate Admissions Office at ASU.

Equipment/Support

The experiment required hardware, software, and some miscellaneous measuring devices/equipment. Each of these are discussed by considering the applicable literature, the experimental requirement, acquisition procedures followed, and the supporting agency.

Hardware. The User Companies A and B cited in the literature employed two primary types of computer terminal displays for data entry, monochrome and color. The color display terminal used was the IBM 3279 with four colors: green, blue, white and red. The function of each color was hardware fixed and could not be manipulated by the operator. Green and blue were for the primary data entry, white was for error messages, and red was used for highlighting actual errors committed. The monochrome display terminal used was not identified except as green characters on black.

A consultant study used only one terminal type, the

ITT Courier 2790 color display computer terminal. It provides the same four colors: green, blue, white and red. The functions of the colors were similarly fixed as stated for the IBM 3279. To simulate a monochrome terminal, a consultant study used the same terminal with the color switch in the "off" position. This caused the initial data entry to be presented in green on black similar to the monochrome display terminals. Error messages were still presented in white, however.

In the reported study, the monochrome display computer terminal used was the ITT Courier 2700. This terminal provides green phosphorous characters on black background and was in current use in the Undergraduate Admissions Office prior to the beginning of the research. The color display used was the ITT Courier 2790, which provides four colors, was the same terminal used in a consultant study. The ITT Courier terminal is functionally similar to the IBM 3279 employed by the referenced User Companies in that both terminals have color displays capable of presenting four colors: green, blue, white, and red. In both terminals, green and blue are used for primary data entry, white is for error messages and red is used for highlighting actual errors. The colors usage was hardware fixed and could not be manipulated by either the

operator or the experimenter. No color display computer terminals were in use at the Undergraduate Admissions Office prior to the experiment. ITT Courier, Phoenix, Arizona, agreed to support the research and provided three of the 2790 color display terminals for a period of one year at no cost to ASU.

Software. Two types of software support were required for the study. The first was software to control the data entry task that was accomplished by the operators. The second was software needed to gather the data required by the research.

The software controlling data entry was currently in use by the operators at the Admissions Office. This software presented the application form on the existing monochrome display computer terminals and controlled operator data entry. It also stored the information entered into the University student data base. Since this software was found to also be compatible with the color display computer terminals, no software changes were necessary. The colors were presented as stated previously.

The software allowing the data collection required by the research was written by the Department of Computer Services at ASU. A formal job request was submitted to

this department which included permission from the Director of Admissions to allow the research in the Undergraduate Admissions Office. The primary contacts for this request were Mr. David Daily and Mr. Mark Burnison, Office of Administrative Systems and Programming.

The job request stipulated the items of data that were necessary. Included for each new application entered were: the date of entry, start time of entry, end time of entry, terminal number on which the entry was accomplished, identification number of the operator entering the application, the applicant's social security number, and any computer recognized errors made by the operator. The date of entry was the calendar day, month, and year in which the operator entered the new application. Start time of entry was operationally defined as that time at which the operator began the session by entering the social security number of the new applicant. The session end time was recorded when the operator stroked the entry or return key on the terminal to allow the computer to accept the information entered for the new applicant. This was done only once per application. The terminal number was a two digit unique number assigned to each of the terminals used at the Undergraduate Admissions Office. Each terminal number was

recognized by the computer since the terminals were hard wired. The identification number of the operator was the user number requested when the operator first signed on to the terminal. The sign on occurred prior to an operator's use of a terminal permitting the operator to use any terminal of the assigned type. The prospective student's social security number was a part of the information entered on each new applicant, which would later become the student's ASU identification number after admission to the University. The final requirement for the data collection software was to report any errors made by the operator during data entry for a new application. The admissions program had an internal error detection routine that looked for two types of errors. The first type was a single field error, which occurred when an error was made in any one of the data entry fields (one example would be the entry of the letter "N" rather than "M" or "F" in the sex field). The other error type was a multiple-field error. The computer software was designed to run comparisons between such fields as zip code and state; if the two did not match an error was generated.

Miscellaneous Measuring Devices/Equipment. In addition to the primary hardware discussed earlier, the study required use of several other measuring devices and

equipment. These included a light meter, a temperature/humidity recorder, a set of plates for testing for potential color blindness of the operators, new application entry logs, and subjective response survey instruments.

A light meter and temperature/humidity recorder were provided and calibrated by the ASU Development Shop. These instruments were used to insure that over the period of data collection the variables of light, temperature, and humidity did not change appreciably.

A set of Dvorine Pseudo-Isochromatic plates for testing the potential color blindness of the operators was provided by the ASU Department of Psychology. Linksz (1964) feels that "as a screening device the Dvorine Test is by far the superior one (over the popular Ishihara Test)" (p. 238). This test was administered to each data entry operator.

A "new application entry log" (Appendix B) was used by each of the operators. This log was designed and supplied to the operators by the experimenter. The purpose of the log was to record computer problems, computer terminal problems and interruptions during the operator sessions. The entries could then be used in the analysis portion of the research to identify invalid data

points. Computer problems included crashes as well as slow downs in response. A computer crash during an operator session caused total loss of any information already entered on that new applicant. A slow down was defined as more than a 15 second wait for the computer to respond to an operator request. This waiting period was measured by the individual operators. One possible computer terminal problem was if a key was hit in error that caused premature termination of the session. Interruptions included interface during a session with other operators, management personnel, or students requesting information. Any interruption of more than 15 seconds as measured by the operator was logged.

The log requested entry of seven items when one of the above instances occurred. The first item was the date of the occurrence. The approximate begin time and end time of the session were the next two items. The fourth item was the terminal number on which the operator was working. The next item was the social security number. The last two required only a check mark in one of the two columns headed "hold" and "abort". Hold indicated that the operator was working on a new applicant entry when one of the occurrences above happened and the operator allowed the information to remain on the screen. This would cause

the additional time not used for data entry to be a part of the overall session time for that applicant. Abort implied that the operator cleared the screen and started over. These items, plus indication at the top of the log of the operator user identification number, made each problem that occurred during the experiment identifiable. All sessions logged were subsequently removed from the research data.

Two surveys were used as subjective measuring devices in the experiment. The first of these was a single terminal evaluation survey (Appendix C) that allowed the operator to rate the terminal used for a particular period of time. This survey was designed to investigate comments, cited in the literature, made by operators who were users of color terminals and/or subjects in research studies concerning data entry via color computer terminals. These comments are reported in the Dependent Variable section of this chapter. They included areas concerning job satisfaction, terminal satisfaction, effects of interruptions, glare, eyestrain, headaches, and physical fatigue. Fourteen questions were asked to allow operator evaluation in these areas. The first eleven questions were answered on a five point scale. The last three questions were open ended to acquire operator's

comments about the terminal.

The second survey was a multiple terminal evaluation (Appendix D). This survey allowed the operators to compare the two terminal types used during the experiment. The instrument was designed to replicate a survey that was used in a consultant study (Appendix E). The survey consisted of six statements to which the operators responded "color", "no difference", or "monochrome" as appropriate. There were also five open ended questions asked as a part of this survey for operator comments about the terminals. The responses to this survey were compared to those of a consultant study.

Experimental Procedures

Once the experimental design was established, procedures to enact the experiment were considered. These procedures involved obtaining consent for the experiment, pretesting, training, data collection, and data analysis. Each of these areas are discussed.

Consent

Consent for accomplishing the experiment was necessary in a number of areas. The first was to acquire consent from the Director of Admissions at ASU to conduct the research in the Undergraduate Admissions Office. Once this was obtained, the necessary forms were filed to gain

consent to perform an experiment involving human subjects. These forms were submitted for approval by the ASU Interdisciplinary Committee on Human Experimentation. The final consent was from those operators participating in the study. A meeting to explain the research requirements was held with the operators. The primary thrust of the research was identified as an evaluation of the two terminal types: monochrome and color. However, information was not provided to the subjects concerning the dependent and independent variables of interest. Specifically, no reference was made to color being the primary independent variable. Although it was realized that the presence of color could not be hidden, to avoid biasing the operators it was not discussed. Each operator was requested to sign a consent form (Appendix F).

Pretesting

Pretesting was accomplished for the research hardware, software, and survey instruments used. Procedures followed to accomplish this are discussed.

Hardware. As has been noted, both of the terminals were manufactured by ITT Courier. The monochrome display computer terminals currently in use for the data entry task in the Undergraduate Admissions Office were evaluated to insure that each of them was operating properly. The

new color display computer terminals were tested by ITT Courier personnel prior to shipping them to ASU, followed by subsequent testing by the Department of Computer Services during installation. All were in proper operating condition during the study.

Software. The software controlling data collection was pretested by the Office of Administrative Systems and Programming. The software was also tested by the researcher for a period of three months prior to the research data collection. This testing involved the entering of new applications by the Undergraduate Admission Office operators in the presence of the researcher. Each item of information that the software was designed to collect was recorded as the application was entered. The start and end times were recorded to the nearest second using a digital watch. All of the recorded items were compared with the software data collection for accuracy. The software was operating properly with all computer recorded times being less than 5% greater than and in most sessions exactly equal to the watch times. These tests were run prior to the research data collection as well as periodically throughout the collection period. In addition to this testing, daily printouts were obtained of the collected information to insure that the software

continued to operate properly.

<u>Surveys.</u> The two survey instruments administered were (1) the single terminal evaluation survey (Appendix C) and (2) the multiple terminal comparison survey (Appendix D). The single terminal evaluation survey was pretested using six operators who were not involved in the study but who had previously accomplished the data entry task using the monochrome display computer terminal. The operators responded to the survey questions separately. Immediately following an operator's completion of the instrument, a private post test interview was held allowing her to evaluate the survey. This interview was necessary to insure proper interpretation of the questions and to collect any suggestions concerning deletion and/or addition of questions. The interviews resulted in an assurance by each of the operators that the questions were clear and concise. No questions were added to or deleted from the instrument as a result of the pretesting. The multiple terminal comparison survey was not pretested because this instrument was adapted from a consultant study and it was desired to compare/contrast the survey responses from these two studies. Hence no changes were possible in the format and no pretesting accomplished.

Training

Training for the operators was considered with respect to accomplishing the data entry task and use of the two terminal types. As the operators had from 1.5 to 3.5 years experience performing the data entry task via the monochrome display computer terminal, no training was given in accomplishing the task or using the monochrome terminal. None of the operators had ever used the color display computer terminal; however, except for the color display this terminal was virtually identical to the monochrome terminal. Therefore no training was required. As part of the analysis, the data were checked for the possible existence of a learning curve with respect to the use of the computer terminal.

Data Collection

Data collected to address the research questions of interest were in two general categories: objective and subjective. The objective category included testing the environmental conditions in which the operators were working, testing the operators for possible color deficiency, data from the new application entry log, and collecting the required information from the data entry task. The subjective category involved the administering of the single terminal evaluation survey and the multiple

terminal comparison survey. The procedures followed in collecting the data in each of these categories are discussed.

Objective. The first type of data collected in the objective category was measurements of the environmental conditions in which the operators were working. The conditions of lighting, temperature, and humidity were measured with the appropriate equipment three times during the experiment: prior to data collection, in the middle of the experimental period, and at the end of the experiment. At each measurement point the lighting was measured in the morning, midday and afternoon of one day. The temperature and humidity were recorded for a minimum of three working days each time they were measured. The environmental conditions were found to be similar at each measurement period throughout the experiment.

Another type of objective data collected was a measurement of operator color vision. The Dvorine Pseudo-Isochromatic Test was administered to each operator at the completion of the experiment. These procedures were followed for two reasons. First, it was felt that to administer the test prior to the experiment would confirm to the operators that color was the key issue in the research and possibly inject bias into the experiment.

Secondly, since only .5% of females are color blind (Demars, 1975) the researcher was virtually assured that all of the operators had normal color vision. This was found to be the case when the test was given following the experiment.

A third form of objective information collected during the experiment was that entered on the new application entry log (Appendix B). An entry was made in this log when either computer problems or other interruptions occurred as described in the Experimental Design section of this chapter. These entries were used to identify erroneous data points gathered by the computer software on the data entry task. These data points were subsequently removed from the research data.

The final type of objective data collection was that gathered by the computer software concerning operator entry of new applicants. This objective data were the primary concern to the research. As the research was interested not only in the effects of color but also in the possible degradation of these effects over time, data were collected for a period of seventeen weeks. The weeks chosen for collection on the data entry task were during the peak months, indicated by historical data as being February to June, of new applicants requesting entry to

ASU. These months were chosen to allow for collection of the maximum possible number of data points during the experiment. To acquire this type of objective data the experiment was split into four phases. Meetings were held prior to each of the phases to explain the data collection period and the operator terminal assignments. Each of these phases are discussed in detail.

Phase 1 served the purpose of establishing a baseline for each operator involved in the study. This phase lasted for four weeks from 7 February to 4 March 1983 with all operators working on the monochrome display computer terminal.

Phase 2 allowed operator performance to be measured when using a monochrome display versus a color display computer terminal for the data entry task. This phase lasted for five weeks from 7 March to 8 April 1983. This phase followed the experimental design of control group and experimental group. The control group, group 1, continued to work on the monochrome display computer terminals. This group consisted of five of the operators. The experimental group, group 2, accomplished the data entry task via the four color display computer terminals. This group consisted of four operators. The operators in each of these two groups were selected to insure that the

two levels of age and two levels of experience with the data entry task were represented in both groups.

Phase 3 data gathering served as verification for any results in phase 2. Phase 3 covered five weeks of data collection from 11 April to 13 May 1983. The terminal assignments for the operator groups discussed for phase 2 were switched. Therefore the five operators of group 1 worked on the color display computer terminals and the four operators comprising group 2 worked on the monochrome display computer terminals. This completed the major data entry task collection period. Another phase, Phase 4, followed during which some special extensions of the research were considered.

Phase 4, the final three weeks of the experiment, involved three special extensions. There were two purposes underlying these extensions. One purpose was to validate some areas of concern in the experimental design. These areas included the possible influence on operator performance of physical differences between the terminals used in this study other than the presence of color. Another area was that of the possible existence of a learning curve in the data and/or change in the effects on operator performance of the independent variables over time. The second purpose of these special extensions was

to emulate as closely as possible a consultant study in an attempt to verify some of their stated results. Each extension is described below.

The first extension involved usage of the computer terminal for a continuous period of two hours. This was done to investigate possible physical terminal differences and to emulate as closely as possible a consultant study. The time requirement was similar to that of a consultant study. This extension involved one operator using one of three terminal configurations for a period of one week each, two continuous hours per day. The operator was relieved of all other office duties for this two hour period. The terminal configurations were monochrome, color, and color with the color switch in the off position. The latter configuration caused the display of the color computer terminal to use only two of the four colors. Data entry appeared in green and any error messages were indicated in white. This configuration was the simulated monochrome configuration used in a consultant study.

The second extension considered a comparison of data entry performance using the color display computer terminal with the color switch off versus using the monochrome display computer terminal. This was done to

investigate the possible physical terminal differences' effects on operator performance and the comment made in a consultant study. The comment made was that "the productivity differences between color and monochrome may be understated as a result of the white error messages providing a color advantage when processing in monochrome" (Shafer, 1982, p. 10). The current study investigated this comment by comparing the data for operators using the color terminal with the switch off to the data for those using the monochrome terminal. For one week two operators used one of these terminals and two used the monochrome. In the second week, these assignments were switched.

The third extension was concerned with the investigation of randomness of the data over an extended period of time. Randomness was defined as the data remaining similar for each operator throughout the collection period. This allowed further investigation of the possible existence of a learning curve and/or change in the effects on operator performance of the independent variables over an extended period of time. For this three week period four operators continued using the terminal they used during the previous five weeks. Two operators used the color display computer terminal and the other two used the monochrome display computer terminal. This

allowed analysis of eight weeks of data collect on four operators.

Subjective. The subjective data collection consisted of two survey instruments: single terminal evaluation survey and multiple terminal comparison. Each instrument was discussed previously in the Miscellaneous Measuring Devices/Equipment section of this chapter. The single terminal evaluation survey (Appendix C) was administered at the end of each of the first three phases of the experiment. The multiple terminal comparison survey (Appendix D) was administered once immediately following completion of the seventeen week experiment. Both instruments were administered to all nine operators at once. Verbal instructions were given concerning proper marking of answers on the survey and confidentiality of the answers given. The instructions at the top of the survey were read aloud. No time limit was imposed. A post survey interview was held as required. This interview was used to clarify any inconsistencies that were noticed by the experimenter from one survey to another for a particular operator. Also discussed at that time were any comments made on the survey that were unclear to the experimenter.

Data Analysis

After all data were collected, rigorous analysis was accomplished by methods suggested by the mathematical literature. The main data from the data entry task were analyzed using two computer software packages. These were the Statistical Analysis System (SAS) and the Biomedical Computer Programs P-Series (BMDP). The data, analysis procedures, results and conclusions based on both the objective and subjective data are described in detail in the following chapters.

IV. DATA EXPLANATION AND PREPARATION

Introduction

The data collected during the seventeen weeks of this research are quite complex. Therefore a short chapter is devoted to explanation and preparation of the data. The explanation will discuss the process by which data were generated as well as describe the actual raw data. The preparations required to format these data for analysis are described. These preparations involved splitting the data into the various phases of the study, reformatting of the data, and elimination of invalid data lines. An example of the raw data and the prepared data is given in Appendix G.

Explanation

A line of raw data was captured by a computer program each time the operator keyed in information concerning a new applicant that was being entered into the system. The computer program was coded by the Office of Administrative Systems and Programming, Arizona State University (ASU) with the cooperation of Mr. Dave Dailey and Mr. Mark Burnison. Each data line was captured when the operator stroked the return or entry key on the computer terminal. Therefore several lines of data were generated for each

new applicant that an operator processed. A new applicant was defined as an application for which no information existed currently on the ASU computer system. Each line of data included the following items. The first item was date as YY/MM/DD. Start time was the second item. In the case of multiple data lines for a particular applicant this time was the same for each line. The end time was the third item and reflects the operator's stroke of the return key on the terminal. Terminal identification by number and operator identification by computer access code were the fourth and fifth items. The ASU student identification number of the applicant was the sixth item. The next items were 70 data fields containing a "0" or a "1" indicating that a particular error had not occurred or had occurred respectively with this applicant entry. If no data entry errors were made, these fields remained blank.

Preparation

In preparing the data for analysis several tasks had to be accomplished. These included splitting the data into the various phases of the study, reformatting the data, and eliminating invalid data lines.

As described previously in Chapter III, Data Collection section, the research involved four phases of

data collection. Phase 1 data were captured with all nine operators using monochrome computer terminals. Phase 2 data represented the experimental group (four operators) using the color display computer terminals and the control group (five operators) continuing to use the monochrome computer terminals. Phase 3 switched these two group terminal assignments. Phase 4 consisted of three special extensions of the research. The data was split into four segments representing these four phases of the study. Each phase of data was stored in a separate file. This allowed for properly addressing the research questions as appropriate to each of the phases.

Reformatting involved a series of operations to prepare the raw data for analysis. The first step was to count the number of errors made by the operator for each applicant entered. The second step was a consolidation of all entries for a particular applicant. This process created one line of data for each applicant entered. It also created the dependent variables of interest to the study which were a total for the number of errors made per applicant entry and the number of seconds taken to enter the applicant into the system. The data set achieved by this reformatting process included the following variables for each new applicant entered as shown in Appendix G:

date (DATE), ASU identification number (ASUID), terminal identification number (TERMID), operator identification number (OPRID), number of errors (COUNT), session time to enter applicant (SESTIME), and time the operator started the entry (BEGTIME). This single line of data totally described each application session.

Elimination of invalid data lines was required in two cases. The first case involved sessions by operators not involved in the study. The second case involved data entry times affected by computer problems or operator distraction and termination. There were nine operators participating in the study. However sometimes due to workload variation, other operators would enter some new applications. The resulting entries caused the capture of data lines which had to be removed. A "new applicant" entry log (Appendix B) was reviewed for possible invalid data entry times. These logs provided the operator means by which to note unusual computer or computer terminal problems and outside interference during the operator session. A computer problem was defined as either a computer crash or more than 15 seconds computer delay incurred during an applicant session. One possible computer terminal problem was if a key was hit in error that caused premature termination of the session. Outside

interference referred to any interruptions of more than 15 seconds. Each of these were described in detail in Chapter III, Miscellaneous Measuring Devices/Equipment section. The new applicant data lines generated during any of the conditions discussed were identified and eliminated from the final data base.

Summary

The data generation and preparation activity resulted in four data files with a total of 6688 lines of data. These data files represented seventeen weeks of processing new applicants by nine operators. A detailed listing of a portion of the data base is in Appendix G. The next step in the research was to examine the data from phase 1 collection period to establish a baseline for each of the operators and to determine the analysis approach to pursue.

V. MONOCHROME TERMINAL BASELINE DATA ANALYSIS

Introduction

Using the data described in the previous chapter, analysis first was accomplished using that portion referred to as phase 1. Phase 1 data collection was for a period of four weeks from 7 February to 4 March 1983. During this time period all nine operators, both the control and experimental groups, worked on the existing monochrome terminals. The purpose for phase 1 was to set a baseline for each operator used in the study. The analysis approach to pursue for the remaining phases would also be determined from these data. A total of 1999 data points were used for this phase of analysis.

There were several questions to be answered using phase 1 data. Are the dependent variables of error count (COUNT) and session time (SESTIME) for each applicant entry related? Is the error count per session different between operators and between the control and experimental groups? Is the daily average of the error count variable random over time for each of the operators? Are the session times different between operators or between the control and experimental groups? Is the daily average of the session time variable random over time for each of the

operators? Each of these questions and related implications is addressed in the following sections of this chapter. Conclusions will complete the discussion.

Dependent Variable Relationship

The first question to resolve using phase 1 data was the relationship between the two dependent variables of interest in the research. These variables were number of errors (COUNT) made by the operator when entering a particular applicant's information and the amount of time (SESTIME) required for this entry. The relationship of these two variables would indicate whether the two dependent variables should be considered independently or combined during further analysis of the data. The SAS correlation procedure was used to address this question; in particular the Pearson product moment correlation coefficient was utilized. This is "the most commonly used method of correlation" (Conover, 1971, p. 244). It is calculated by dividing the sample covariance by the product of the two sample standard deviations (Conover, 1971).

> This measure of correlation may be used with any data of a numeric nature without any requirements concerning the scale of measurement or the type of underlying distribution. It meets the necessary requirements of an acceptable measure of correlation (Conover, 1971, p. 245).

The Pearson product moment correlation coefficient was calculated for each operator, for the two groups (experimental and control) of operators, and for the entire phase 1 data set. Table 5.1, Correlation Coefficients for Various Data Separations, lists the results of the correlation procedure for these three separations of the data. The first column of the table specifies the applicable data separation. The column also includes a number identifying which operator's session data was under consideration for the operator separation and a number indicating the control group, group 1, and the experimental group, group 2, for the group separation. The second column lists the number of data points used for each of the correlations between COUNT and SESTIME. The next column lists the Pearson product moment correlation coefficient (R) values. The final column lists the p values associated with testing the null hypothesis that the correlation coefficient is zero.

The Pearson product moment correlation coefficient, column 3, for the relationship between the COUNT and SESTIME variables was <.2 for all but two of the operators. When considered by groups, experimental and control, the R values were both <.13. The value of R using the entire phase 1 data set was .11.

Table 5.1

CORRELATION COEFFICIENTS

for

VARIOUS DATA SEPARATIONS

Ho: $\rho = 0$
Reject if $p \leq .05$

Separation	Number of Sessions	Pearson (R)	p Value
Operator Number			
1	311	.11	.04*
2	229	.02	.70
3	275	.11	.08
4	337	.10	.06
5	176	.24	.01*
6	225	.12	.08
7	174	.18	.02*
8	113	.08	.38
9	159	.23	.01*
Group			
1	1083	.13	.0001*
2	916	.10	.0015*
Total Data Set	1999	.11	.0001*

*Rejected at assumed significance level of .05

The p values, column 4, indicate that in some cases the null hypothesis that the correlation coefficients just discussed were equal to zero was rejected, and in other cases the conclusion was failure to reject. If an alpha level of .05 was desired, this null hypothesis could not be rejected for operator number 2, 3, 4, 6 and 8. In these cases the p value was >.05. This implies that "any correlation in the sample is primarily the result of chance" (Lewis and Ford, 1983, p. 96). For the other operators, the two groups, and the overall data set the p values were small, <.05, thus rejecting the null hypothesis. This implies that the correlation coefficient was statistically significant.

> It should be remembered that statistical significance does not imply importance. It simply means that the relationship found in the sample is present in the population. Given statistical significance it is then up to the researcher to decide whether the relationship indicated by the correlation coefficient has meaning (Lewis and Ford, 1983, p. 96).

According to Lewis and Ford (1983) a correlation coefficient between 0 and .3 "indicates a weak relationship"(p. 96). Therefore, since the dependent variables of error count and session time have no more than a weak relationship, R=.11, further analysis can consider the two dependent variables separately. Hence the analysis discussion to follow considers COUNT first

and then SESTIME.

Error Count Analysis

Introduction

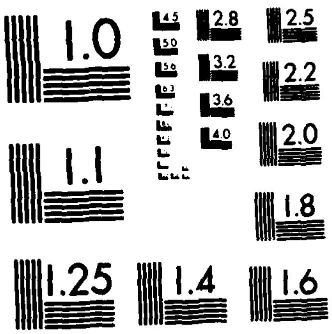
A count of the number of errors (COUNT) was generated for each operator session. Analysis of this dependent variable using the 1999 phase 1 data values was accomplished to answer several questions. These questions required resolution prior to analysis of the data in the later phases of the research. First, is the COUNT different between operators and/or between groups? If this error count variable shows significant differences then a correction factor may be necessary prior to further analysis. Second, are the data with respect to the COUNT variable random over time? Variation in time might indicate a relationship between the dependent variable of COUNT and the independent variable of date of the session. The approach to each of these questions is discussed separately.

Operator and Group Comparison

The possible difference of the error count variable between operators and/or between the control and experimental groups was investigated using a one way analysis of variance. Two null hypotheses were tested. The first of these was that the COUNT variable means for

AD-A145 568 ANALYSIS OF DATA ENTRY PERFORMANCE: CHROMATIC VERSUS MONOCHROMATIC(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH R L ROSE MAY 84 2/3

UNCLASSIFIED AFIT/CI/NR-84-48D F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

the operators were equal. The second that the COUNT variable means for the groups were equal. The ANOVA calculated an F=1.57 with 8,1990 degrees of freedom which indicates insufficient evidence to reject the null hypothesis with respect to operators at p=.13. For the two groups, failure to reject the null hypothesis was also the conclusion with p=.27 resulting from an F=1.23 with 8,1997 degrees of freedom. To further support these conclusions a nonparametric Kruskal-Wallis test was used on each of the above. In both cases failure to reject the null hypothesis was the conclusion using this test. For operators, F=1.18 with 8,1990 degrees of freedom resulted in p=.31. For groups, F=.65 with 1,1997 degrees of freedom resulted in p=.42. These statistics imply that there is considerable risk in concluding that the error count means are different between operators or groups of operators and corrections for any differences should not be made. Hence no correction factor was applied prior to the analysis using this variable in the future phases.

Randomness of Errors Over Time

The second question of interest using the phase 1 data was whether the error count variable (COUNT) was random over time, date of the session (DATE), for each of the operators. Any trends, lulls, or peaks would be of

interest as this would indicate relationship between these two variables. Such a relationship in the data might be explained by the presence of a learning or some other undesirable experimental affect. If such an affect were found to be present, it would be removed by applying the appropriate procedure. To investigate the randomness of the COUNT data over time (DATE), regression analysis was used. The steps followed were to plot the data allowing visual inspection of the possible relationship, to estimate the slope in the linear regression model and test its significance, and to estimate the second order coefficient in the quadratic model and test its significance (Lewis and Ford, 1983). The t-test was used to test the significance of the model coefficients (Lewis and Ford, 1983). The null hypothesis being considered was that the coefficient was zero.

The results of these tests are delineated in Table 5.2, Phase 1 Data Randomness of COUNT vs DATE. The first column identifies the operator for which the results apply. The second column lists the p value for the test of significance of the slope in the linear model. The final column presents the p value for the test of significance of the second order coefficient in the quadratic model. Assuming an alpha value of .05, the

Table 5.2

PHASE I DATA RANDOMNESS of COUNT vs DATE

Ho: Model Coefficient = 0

Reject if $p \leq .05$

Regression Model / Operator Number	First Order Slope p Value	Second Order Coefficient p Value
1	.64	.06
2	.38	.31
3	.61	.68
4	.46	.23
5	.67	.38
6	.68	.59
7	.24	.96
8	.01*	.62
9	.88	.24

* Rejected at assumed significance level of .05

results of the test for all but operator 8 indicated considerable risk in assuming that the variables of COUNT and DATE were related. Hence the data were assumed to occur at random over time. Since for operator 8 the conclusion was to reject the null hypothesis for the slope value, this question of randomness over time was considered in each of the following phases of the study as operators are introduced to the color terminals.

Session Time Analysis

Introduction

Total time for a session was generated for each operator entry of each new application. Analysis of phase 1 data with respect to this dependent variable of session time (SESTIME) was accomplished to answer several research questions. These answers are pertinent to the method of further analysis. First, is SESTIME different between operators and/or between the control and experimental groups? If the operators involved in this study worked at significantly different rates of speed on the same terminals, then a correction factor would be applied to future data prior to analysis. Second, is the data with respect to the SESTIME variable random over time for each of the operators? Variation in time might indicate a relationship between the dependent variable of session

time and the independent variable of date of the session. Each of these questions is discussed separately.

Operator and Group Comparison

The question of difference between the session times (SESTIME) for the operators working on the same terminals was approached using a one way analysis of variance. This analysis was accomplished both for individual operators and for the two groups, control and experimental, of operators. Two null hypotheses were tested. The first of these was that the SESTIME variable means for the operators were equal. The second that the SESTIME variable means for the groups were equal. Both of these null hypotheses were rejected with p=.0001. The ANOVA values were F=41.95 with 8,1990 degrees of freedom and F=45.33 with 1,1997 degrees of freedom respectively. The p value was the same for both the analysis of variance and the nonparametric Kruskal-Wallis test. The table values for this latter test were F=62.63 with 8,1990 degrees of freedom and F=40.73 with 1,1997 degrees of freedom for the two hypotheses respectively. It was concluded that the operators worked at significantly different rates of speed when accomplishing the same task using the same computer terminal. To correct for this difference between operators a weight factor was calculated to equalize the

session times of each of the operators. This weight factor was defined as the ratio of the operator mean to the grand mean for the entire data set (103.48). Table 5.3, Operator Mean Session Time and Weight Factor, shows these calculated values for each operator. The first column of the table lists the operator number. The second column lists the average session time for each operator during this phase of data collections. The weight factor for each operator is listed in column three. Each of the session times was divided by this weight factor to arrive at the corrected session time (CSESTIME) variable to be used in future analysis.

Randomness of Session Time Over Time

The second question of interest with respect to the dependent variable SESTIME was whether the data were random over time for each of the operators. Analysis to answer this question might identify the presence or absence of a possible learning and/or other undesirable experimental affect. To investigate this possibility regression analysis was accomplished following the same procedures as for error count analysis previously described in the Randomness of Errors Over Time section of this chapter. The variables used in this analysis was corrected session time (CSESTIME) versus date of data

Table 5.3

OPERATOR MEAN SESSION TIME AND WEIGHT FACTOR

Operator Number	Mean Time (seconds)	Weight Factor
1	97.27	.94
2	84.32	.81
3	92.87	.90
4	112.73	1.09
5	133.56	1.29
6	106.98	1.03
7	111.39	1.08
8	83.00	.80
9	109.63	1.06

entry (DATE). The results of the test for significance of the slope in the linear regression model and the second order coefficient in the quadratic model are presented in Table 5.4, Phase 1 Data Randomness of CSESTIME vs DATE. The information in this table is presentea similarly to that of Table 5.2 previously discussed. The null hypothesis tested was that the coefficients were zero. Assuming an alpha value of .05, the results for all but operator 6 indicated considerable risk in assuming that the variables of CSESTIME and DATE were related. Hence the data were assumed to occur at random over time. Due to the fact that for the slope value for the data of operator 6 the conclusion of the significance test was to reject the null hypothesis, this question was considered in each of the following phases of the study.

Conclusions

The phase 1 data analysis answered a number of questions of concern to the research. Resolution of these questions was necessary to allow for proper analysis of the data in the future phases of the study. No significant relationship was found to exist between the dependent variables of error count and session time. Therefore future analysis considered the two dependent variables separately.

Table 5.4

PHASE 1 DATA RANDOMNESS of CSESTIME vs DATE

Ho: Model Coefficient = 0

Reject if $p \leq .05$

Regression Model / Operator Number	First Order Slope p Value	Second Order Coefficient p Value
1	.60	.47
2	.33	.94
3	.96	.09
4	.43	.06
5	.38	.60
6	.00*	.09
7	.70	.25
8	.35	.13
9	.88	.06

* Rejected at assumed significance level of .05

There was found to be considerable risk in concluding that the error count means were different between operators or between the two groups of operators. Therefore no correction factor was applied in analysis of the other data phases. Also the average daily error count was indicated to occur at random over time for all but one of the operators. Therefore this question was considered when analyzing the other phases of the experiment.

The session time dependent variable was found to be significantly different between operators and between groups. This led to the calculation of a weight factor and the creation of the corrected session time variable. The corrected values were used in all future analysis. The average daily corrected session time was indicated to occur at random over time for all but one of the operators. Therefore this question was examined for each of the following phases of the experiment.

Analysis of the data in the following phases of the research was accomplished on the error count and the session time dependent variables separately. The findings with respect to each of these variables in the areas of operator and group comparison and randomness of the data over time were applied in accomplishing the color versus monochrome display data analysis.

VI. COLOR VS MONOCHROME TERMINAL DATA ANALYSIS

Introduction

Once the required baseline questions were answered using phase 1 data, the analysis considered the research questions involving color vs monochrome display computer terminal usage for data entry. The data for this analysis were collected in two phases, phase 2 and phase 3.

Phase 2 allowed operator performance to be measured when using a color versus a monochrome display computer terminal. This phase of data collection was for a period of five weeks from 7 March to 8 April 1983. During this time period the operators were split into two groups. Group 1, the control group, consisted of five operators who continued to work on the monochrome terminals. Group 2, the experimental group, consisted of four operators who accomplished the data entry task via the four color computer terminal. A total of 2185 data points were collected during phase 2.

Phase 3 data collection served as verification for any results in phase 2. This phase also covered a five week period from 11 April to 13 May 1983. The two group assignments to terminals for the data entry task were switched during this phase. Group 1 operators were now

assigned to the color computer terminals and group 2 to the monochrome computer terminals. A total of 1811 data points were collected during phase 3.

The analyses of these two phases of data collection were performed similarly and are presented simultaneously to allow comparison of results. These analyses are discussed here in four sections: preliminary questions, error count analysis, corrected session time analysis and conclusions.

Preliminary Questions

Two preliminary questions were answered prior to the research analysis of the data. The first question addressed data randomness over time. The second question concerned the analysis model and technique. Each of these questions are discussed.

Data Randomness Over Time

Randomness of the data over time was of concern for each operator due to the possible presence of a relationship between the error count per session (COUNT) variable versus the date of data entry (DATE) and/or the corrected session time (CSESTIME) variable versus DATE. The relationship could indicate the existence of a learning affect caused by the color computer terminal introduction and/or some other undesirable experimental

affect. If the data indicated no trends over time then the assumption would be made that these affects were not present. These phenomena were investigated in both phase 2 and phase 3 data using regression analysis. As discussed in the Randomness of Errors Over Time section of Chapter IV, the data were plotted allowing for visual inspection of the possible relationship, then, the slope in the linear regression model estimated and tested for significance, and finally, the second order coefficient estimated in the quadratic model and tested for significance (Lewis and Ford, 1983).

The plots of COUNT vs DATE and CSESTIME vs DATE showed no obvious trends for either phase 2 or phase 3 data. The slope and second order coefficients were calculated and their significance checked with a t-test using the linear and polynomial regression procedures available in SAS. The null hypothesis tested was that these model coefficients equaled zero. The results of these tests are presented in Table 6.1, Phase 2 Data Randomness of COUNT vs DATE, Table 6.2, Phase 3 Data Randomness of COUNT vs DATE, Table 6.3, Phase 2 Data Randomness of CSESTIME vs DATE and Table 6.4, Phase 3 Data Randomness of CSESTIME vs DATE. The first column in each of these four tables identifies the operator for which the

Table 6.1

PHASE 2 DATA RANDOMNESS of COUNT vs DATE

Ho: Model Coefficient = 0

Reject if $p \leq .05$

Regression Model / Operator Number	First Order Slope p Value	Second Order Coefficient p Value
1	.51	.33
2	.77	.31
3	.29	.65
4	.43	.18
5	.73	.61
6	.15	.88
7	.24	.77
8	.26	.07
9	.92	.45

* Rejected at assumed significance level of .05

Table 6.2

PHASE 3 DATA RANDOMNESS of COUNT vs DATE

Ho: Model Coefficient = 0

Reject if $p \leq .05$

Regression Model / Operator Number	First Order Slope p Value	Second Order Coefficient p Value
1	.27	.55
2	.19	.90
3	.30	.56
4	.31	.75
5	.18	.38
6	.83	.07
7	.62	.96
8	.31	.32
9	.20	.06

* Rejected at assumed significance level of .05

Table 6.3

PHASE 2 DATA RANDOMNESS of CSESTIME vs DATE

Ho: Model Coefficient = 0

Reject if $p \leq .05$

Regression Model / Operator Number	First Order Slope p Value	Second Order Coefficient p Value
1	.81	.92
2	.61	.46
3	.10	.57
4	.18	.40
5	.10	.21
6	.22	.79
7	.82	.90
8	.81	.43
9	.91	.51

* Rejected at assumed significance level of .05

Table 6.4

PHASE 3 DATA RANDOMNESS of CSESTIME vs DATE

Ho: Model Coefficient = 0

Reject if $p \leq .05$

Regression Model / Operator Number	First Order Slope p Value	Second Order Coefficient p Value
1	.69	.60
2	.11	.39
3	.88	.22
4	.43	.41
5	.38	.92
6	.88	.11
7	.14	.91
8	.70	.59
9	.56	.96

* Rejected at assumed significance level of .05

results apply. The second column lists the p value for the test of significance of the slope in the linear regression model. The final column presents the p value for the test of significance of the second order coefficient in the quadratic model. The values of p in these last two columns indicate whether or not to reject the null hypothesis. If p is less than or equal the desired value of alpha, then rejection of the null hypothesis is concluded. Assuming an alpha value of .05, the results of the tests indicated failure to reject the null hypothesis under consideration for this part of the analysis. All values of p are greater than .05. This implies considerable risk in assuming that COUNT vs DATE or CSESTIME vs DATE have either a linear or quadratic relationship in either phase 2 or phase 3 of the data. Therefore the assumption was made that no learning affects or other undesirable experimental affects existed in either phase of the data and no further corrections were made to the data.

Analysis Technique and Approach

The analysis technique planned for use in answering the primary research questions was analysis of variance (ANOVA). This technique allows for analysis of several independent variables simultaneously. It was desired to

consider as many of these variables simultaneously as possible in the ANOVA model in order to analyze all possible interactions. In considering a 4-factor analysis of variance model including all four independent variables, a majority of the cells were empty. Therefore a 3-factor ANOVA model was considered. When investigating the required assumptions for use of the ANOVA technique, the 3-factor models for analyses were found to be infeasible. As none of interaction effects were found to be statistically significant for the 3-factor models, 2-factor ANOVA models were considered. The generic form of the 2-factor model considered for analyses was:

$$y(ijk)=u + \tau(i) + \beta(j) + \tau\beta(ij) + \epsilon(ijk)$$

i=Treatment Level
j=Characteristic Level
k=Observation Number

where y(ijk) was the (ijk)th performance (corrected session time or error count) observation and u was the overall mean performance effect. The $\tau(i)$ represented the true effect of the ith level of terminal treatment (color or monochrome). The $\beta(j)$ was the true effect of the jth

level of the characteristic factor (age, experience level or time of day of data entry). $(\tau\beta)(ij)$ represented the effect of the interaction between the ith terminal treatment and the jth characteristic factor and $\epsilon(ijk)$ was the random error component of the (ijk)th observation. Table 6.5, Variables for Analysis, lists the independent variables with their associated levels and the dependent variables for each term defined in the models used for analyses. The 2-factor models of interest were: TERMGP x AGEGP, TERMGP x EXPLV, and TERMGP x TOD. These were analyzed first for all forms of the dependent variable of error count and then for the dependent variable of corrected session time.

Error Count Analysis

Introduction

There were four research questions of concern with respect to the dependent variable of error count. Does color display usage affect input error rate? Are the effects of color display usage related to the age of the operator? Are the effects of color display usage related to the experience level of the operator? Are the effects of color display usage related to the time of day of data entry? These research questions were considered by separate analysis of phase 2 (group 1 using monochrome

Table 6.5

VARIABLES for ANALYSIS

INDEPENDENT VARIABLES

<u>Treatment Variables $\tau(i)$</u>

Terminal Group (TERMGP)
- Color
- Monochrome

<u>Characteristic Variables $\beta(j)$</u>

Age Group of Operators (AGEGP)
- Less than or equal to 35 years
- Over 35 years

Experience Level of Operators (EXPLV)
- Less than or equal to 2 years
- Over 2 years

Time of Day of Data Entry (TOD)
- Entries prior to noon
- Entries at or after noon

DEPENDENT VARIABLES $y(ijk)$

<u>Error Rate</u>
- Error Count (COUNT)
- Error count Rank Transformation (RCOUNT)
- Error Count Ratio (CNTRAT)
- Error Count Ratio Rank Transformation (RCNTRAT)

<u>Session Time</u>
- Corrected Session Time (CSESTIME)
- Ranked Corrected Session Time (RCTIME)

terminal, group 2 using color terminal) and phase 3 (group 1 using color terminal, group 2 using monochrome) data. The analyses are discussed simultaneously for both phases. Initially data were checked as to whether or not the required assumptions for the ANOVA technique were supported. This is discussed first. The research questions are then resolved using the appropriate models.

ANOVA Assumptions

The analysis of variance technique makes four basic assumptions about the data to be analyzed. These are generally referred to as independence, normality, homogeneity of variances and additivity (Berenson, Levine, and Goldstein, 1983). A description, the analytical technique employed to test, the results and any changes needed to be made to the data or the models are discussed for each of the four assumptions with respect to the ANOVA models under scrutiny.

Independence. The assumption of independence implies that the observed value in any cell of the ANOVA model has no effect or influence on any other observed values in that cell or any of the other cells (Berenson et al., 1983). This required assumption was met due to several factors inherent in the experiment. The operators worked separately on their own unique set of data entry forms.

Also, data were collected only on new applicants and once the initial entry was made by an operator, this applicant was no longer considered new. Hence two data points could not be collected on the same application. Therefore it is assumed for this research that the data were independent.

Normality. Normality of the residuals in each cell of each ANOVA model was considered. Since the operators were experienced with the tested data entry task, they made very few errors. Therefore most of the residuals were highly skewed to the left, hence the normality assumption was not satisfied. A new variable was created in attempt to more closely meet this assumption. The variable was defined as the ratio of the sum of the number of errors per day for each operator over the number of applications entered by the operator for that day. This variable was referred to as the error count ratio. The creation of the variable allowed consideration of two 2-factor ANOVA models. These were terminal type (TERMGP) x age group of the operator (AGEGP) and terminal type (TERMGP) x experience level group of the operator (EXPLV). The normality of the residuals were considered for each of these models and compared to the original variable of error count to detect any improvement with respect to this assumption. The Kolmogorov D-statistic procedure in SAS

was used to test the normality of the residuals.

Neither the error count (COUNT) nor the error count ratio (CNTRAT) variables using either phase 2 or phase 3 data passed this test with an assumed alpha value of .05. However, the fixed effects ANOVA model is relatively robust against departures from normality provided the departure is not of extreme form, that is the distribution of residuals are not highly skewed but bell-shaped (Neter and Wasserman, 1974). The coefficients of skewness and kurtosis were calculated to investigate if the residuals formed a bell-shaped curve. These coefficients measure the symmetry and the peakedness respectively of the distribution. If the magnitude of the coefficient of skewness was less than 1.4 and the coefficient of kurtosis was less than 2.5, the distribution of the residuals was assumed to be close to bell-shaped (Lewis and Ford, 1983). Under these conditions departure from normality was assumed not to be of extreme form. Table 6.6, Departure from Normality for the Error Variable lists the coefficients of skewness and kurtosis for several forms of the error variable: COUNT, CNTRAT, RCOUNT AND RCNTRAT. The first column of the table lists the variable form to which the values apply. Each of these forms are discussed. The next two columns delineate the

Table 6.6

DEPARTURE from NORMALITY for the ERROR VARIABLE

Phase 2 (P2) and Phase 3 (P3)

Model / Variable	TERMGP x AGEGP Skewness (P2/P3)	TERMGP x AGEGP Kurtosis (P2/P3)	TERMGP x EXPLV Skewness (P2/P3)	TERMGP x EXPLV Kurtosis (P2/P3)
COUNT	7.0/7.5	59.7/67.3	7.2/7.5	61.3/67.4
CNTRAT	2.2/3.1	8.6/14.1	2.7/3.3	9.6/16.3
RCOUNT	3.4/3.2	9.5/ 8.4	3.4/3.2	9.6/ 8.4
RCNTRAT	.3/ .5	-1.5/-1.4	.4/ .5	-1.4/-1.4

coefficients of skewness and kurtosis for the TERMGP x AGEGP model first using phase 2 (P2) and then phase 3 (P3) data. The values of these coefficients for each of the phases are separated by a slash (/). The last two columns list these same values for the TERMGP x EXPLV model.

For the error count (COUNT) variable, the first line of data in Table 6.6, the coefficients of skewness were greater than 1.4 and the coefficients of kurtosis were greater than 2.5 for both phases of data and both models. The assumption made was that the departure from normality of the residuals for the COUNT variable was of extreme form. The coefficients of skewness and kurtosis were then calculated for the error count ratio (CNTRAT) variable. These values are presented in line 2 of Table 6.6. As with the COUNT variable the coefficients of skewness were greater than 1.4 and the coefficients of kurtosis were greater than 2.5 for both phases of data in both models. These values were all smaller than the respective values for the COUNT variable, but still indicated a departure from normality of extreme form. To alleviate this departure, the literature suggests use of a transformation of the data.

The transformation investigated was a rank transformation, a "robust procedure that appears to behave

remarkably well" (Conover and Iman, 1976, p. 1357) in improving departures from normality of extreme form.

> The idea of the rank transform is simple. If there is a parametric method available for analysis of the data, but the assumptions of the parametric method are not appropriate for the data, then one merely replaces the data with their ranks, ranking everything from smallest to largest. Then the parametric method of analysis is applied to the ranks rather than the original data. The idea of replacing the data with the ranks is to transform the original observations into numbers that more nearly satisfy the assumptions of the parametric model, and at the same time retain all of the ordinal information contained in the original data. (Conover and Iman, 1976, p. 1356)

This rank transformation was accomplished for both the error count variable and the error count ratio variable. As with the COUNT and the CNTRAT variable, neither the error count rank transformation (RCOUNT) variable nor the error count ratio rank transformation (RCNTRAT) variable passed the Kolmogorov D-statistic test for normality. The null hypothesis that the cell residuals were distributed normally was rejected for both phase 2 and phase 3 data. The coefficients of skewness and kurtosis of the transformed data were investigated to assure that this departure from normality was not of extreme form. As noted by the values listed in line 3 of Table 6.6, the RCOUNT variable's data did not meet the bell-shaped criteria. However the rank transformation of the error

count ratio (RCNTRAT) did based on the values listed in line 4 of Table 6.6. Hence, for the RCNTRAT variable it was assumed the departure from normality was not of extreme form, thus the normality criterion was satisfied.

Homogeneity of Variances. The third assumption required for use of the ANOVA technique was equality of variances of the residuals across the cells of the ANOVA table. "In practice, lack of normality and unequal variances tend to go hand in hand" (Neter and Wasserman, 1974, p. 105). Further, the transformation which helps in making the distribution of the residuals more normal also is effective in correcting the lack of equality of variances (Neter and Wasserman, 1974). To insure this was the case for the current research data, equality of variances was investigated for all four forms of the error variable considered in the previous Normality discussion.

Several methods are available to determine if the variances are equal. Two of the most popular are the Bartlett and the Hartley tests. Though these two are widely used, both are very sensitive to the normality criterion. The Hartley test also requires that the cell sizes be equal which was not the case in the current research. The modified Levene's L test is robust to the normality assumption and does not require equal cell sizes

(Berenson et al., 1983). Therefore, Levene's L test was used to investigate equality of variances of the cells of both the TERMGP x AGEGP and the TERMGP x EXPLV ANOVA models.

The results of Levene's L test are presented in Table 6.7, Homogeneity of Variances for the Error Variable. The first column of the table lists the form of the error variable for which the information on that line pertains. The second and third columns present the p values for the test of equality of variance of the cells of the ANOVA models of TERMGP x AGEGP and TERMGP x EXPLV respectively. These values are presented for both phase 2 and phase 3 data, separated by a slash. Assuming an alpha value of .05, any p value less than or equal to .05 indicates the data supports rejection of the null hypothesis that the variances are equal. There were only three instances where the data supported equal variances. For these the p value was greater than .05 indicating considerable risk in assuming that the cell variances were not equal. For the error count ratio (CNTRAT) variable, TERMGP x AGEGP, phase 3 data, the p value was .36. For the ranked transformation (RCNTRAT) of this variable both of the models for phase 3 resulted in p values greater than .05 (.73 and .77). Hence the assumption was made that the

Table 6.7

HOMOGENEITY of VARIANCES for the ERROR VARIABLE

Phase 2 (P2) and Phase 3 (P3)

Ho: Cell Variances Are Equal

Reject if $p \leq .05$

Model / Variable	TERMGP x AGEGP p Value (P2/P3)	TERMGP x EXPLV p Value (P2/P3)
COUNT	.00*/.00*	.00*/.01*
CNTRAT	.00*/.01*	.00*/.36
RCOUNT	.00*/.00*	.00*/.01*
RCNTRAT	.01*/.73	.02*/.77

* Rejected at assumed significance level of .05

cell variances were equal in these instances. For phase 2 data, the RCNTRAT variable was an improvement with respect to equality of cell variances over the other forms of the variable. Therefore, this form was the one used in further analysis. Rejection of the null hypothesis that the variances were equal was still supported. Provided that the additivity assumption was true, the ANOVA technique was planned in these cases, however the Welch technique was also planned to support the results. The Welch technique is similar to the ANOVA technique except that it computes a W statistic based on the individual cell variances and is therefore applicable when cell variances are unequal (Brown and Forsythe, 1974).

Additivity. Additivity was the final criterion investigated. The ANOVA model yields sufficient statistics that estimate the main factor parameters only if the main factors are additive. The property of additivity was tested by analyzing the interaction terms in each model under scrutiny (Montgomery, 1976; Neter and Wasserman, 1974). The actual ANOVA table is discussed below in Results and Conclusions. Two methods were used based on the degree of satisfaction of the criterion of homogeneity of variances.

For those models whose cells satisfied the

homogeneity of variance criterion, the interaction term was tested directly for significance at the .05 significance level. This method was applicable for both models using phase 3 data. At this level of significance, both the TERMGP x AGEGP and the TERMGP x EXPLV models indicated considerable risk in assuming the interaction effect was significant. Hence it was assumed that the main factors of these models for phase 3 data were additive.

For those models whose cell's data supported unequal variances and thus the Welch technique planned for main effects analysis, the ANOVA interaction statistics were analyzed but with a more stringent criterion. This method was used for two reasons. First, the Welch technique available only tested main effects, thus did not yield statistics on the interaction terms. Secondly, Brown and Forsythe (1974) reported as much as 17% fluctuation in the F statistics derived from the ANOVA when the variances were extreme and not equal. To compensate for a possible fluctuation in the F statistic for the interaction term in the models whose cell variances were not equal, the F statistics were increased by 17% and the p value recalculated. The level of significance was once again considered at .05. This method was required for both

models in analysis of phase 2 data. Both the TERMGP x AGEGP and the TERMGP x EXPLV models' recalculated p values indicated considerable risk in assuming the interaction terms were significant. Therefore it was assumed that the main factors of these models were additive.

Conclusions. Independence was applicable for all models considered for both phase 2 and phase 3 data. Due to the findings with respect to normality of the residuals and homogeneity of variances, the error count ratio rank transformation (RCNTRAT) dependent variable was selected for further analysis. Normality of the residuals for this variable was indicated for both models of each phase's data. Homogeneity of variances for the cell's data of the ANOVA models was supported by the phase 3 data but not for the phase 2 data. The Welch technique was considered in the cases where homogeneity of variance could not be assumed. The assumption of additivity was investigated using the two methods described in that section. Based on the results from this analysis, it was assumed that the main factors of all models under consideration were additive. Hence, the statistics derived from the ANOVA and/or Welch techniques were employed to analyze the 2-factor models whose measured variable was RCNTRAT. The results of analysis are presented next for phase 2 and

phase 3 data.

Results and Conclusions

The statistics derived from the ANOVA and/or Welch techniques applied to the two models under consideration were analyzed to investigate the effect color had on operator input error rate. The models were TERMGP x AGEGP vs RCNTRAT and TERMGP x EXPLV vs RCNTRAT. Results and conclusions based on the analyses of these models' estimated parameters are presented under the appropriate research question. Table 6.8, ANOVA and Welch Results, details the analytical findings. The table lists the calculated values for each model for the ANOVA and/or Welch techniques. For the ANOVA, these include the mean sum of squares (Mean SS), the degrees of freedom (df), the F value (F) and the associated p-value (p). When the Welch technique was applied, the table shows the degrees of freedom associated with the W statistic, the W statistic, and the p value. These values are presented for phase 2 and phase 3 data analyses separated by a slash (/). All hypotheses were tested at the .05 significance level. That is, the null hypothesis of nonsignificance of the factor was rejected if the p value was less than or equal to .05. The results and conclusions are discussed for each of the research questions.

Table 6.8

ANOVA and WELCH RESULTS

RCNTRAT DEPENDENT VARIABLE

Phase 2 (P2) and Phase 3 (P3)

Ho: Factor equal zero

Reject if p≤ .05

Values / Model Factors	Mean SS (P2/P3)	df (P2/P3)	F or W (P2/P3)	p (P2/P3)
TERMGP & AGEGP				
ANOVA				
TERMGP	9649.18/2555.23	1/1	3.39/1.02	.07/.31
INTERACTION	6910.45/ 86.11	1/1	2.43/0.03	.12/.85
ERROR	2844.24/2511.59	195/184		
WELCH				
MAIN EFFECTS		3,73/N/A	1.20/N/A	.31/N/A
TERMGP & EXPLV				
ANOVA				
TERMGP	4807.27/ 124.43	1/1	1.70/0.05	.19/.82
INTERACTION	5305.32/3192.58	1/1	1.87/1.26	.17/.26
ERROR	2836.13/2529.19	195/184		
WELCH				
MAIN EFFECTS		3,76/N/A	2.46/N/A	.07/N/A

* Rejected at assumed significance level of .05

Color vs Error Rate. Does color display terminal usage affect operator input error rate during the accomplishment of a data entry task? The main factor TERMGP was analyzed for both models using each phase of data. The levels of TERMGP were color and monochrome display terminal. The resulting p values are presented in Table 6.8. For the TERMGP x AGEGP model the values were .07 and .31 for phase 2 (P2) and phase 3 (P3) data respectively. The TERMGP x EXPLV model resulted in p values of .19 and .82 for the two phases of data. Since the phase 2 data did not pass the Levene's L test discussed earlier for homogeneity of variances, the Welch technique was applied and the p values are presented only for this phase in Table 6.8. This technique supported the findings of the ANOVA technique. These results consistently indicate that the data do not support that the use of a color display terminal for data entry affected the operator's input error rate significantly different from the use of a monochrome display terminal. Even though the variability due to age or experience level were accounted for, color had no effect.

Color x Age vs Error Rate. Is the effect of color display terminal usage on operator performance as measured by error rate significantly different for particular

levels of operator age group? The interaction factor in the TERMGP x AGEGP model for phase 2 and phase 3 data was used to investigate this question. The levels of TERMGP were color and monochrome. The levels of AGEGP were the younger group of operators (35 years of age or less) and the older group of operators (greater than 35 years of age). The results of testing the null hypothesis that the interaction factor means were equal are presented in Table 6.8. For phase 2 data, the p value was .12. Even if the F value was changed by 17 per cent due to the unequal variances as discussed in the Additivity section of Error Count Analysis, considerable risk was indicated in assuming that the means are significantly different. The p value using phase 3 data of .85 strongly supports this same result. Hence, the data contained insufficient evidence to conclude that color display usage significantly changes operator error rate over monochrome display usage for either the younger or the older level of operators.

Color x Experience vs Error Rate. Is the effect of color display terminal usage on operator performance as measured by error rate significantly different for particular levels of operator experience? The interaction factor in the TERMGP x EXPLV model for phase 2 and phase 3

data was used to investigate this question. The levels of EXPLV were the less experienced operators (2 years or less experience) and the more experienced operators (more than two years experience). The experience referred to was operator on the job experience accomplishing the data entry task used in the research. The minimum operator experience at the beginning of the data collection period was 1.5 years. The p value for phase 2 data was .17 (Table 6.8) and even considering the possible F value fluctuation due to unequal variances considerable risk was indicated in assuming that the means are significantly different. The p value of .26 for phase 3 data supports this result. Hence, the data contain insufficient evidence to conclude that color display usage significantly changes operator error rate over monochrome display usage for either the less or the more experienced group of operators.

Corrected Session Time Analysis

Introduction

The other objective measure of operator performance in this research besides error rate was time required to accomplish the data entry task. There were four research questions of concern with respect to the corrected session time (CSESTIME) dependent variable. Does color display

usage affect the time required for data entry? Are the effects of color display usage related to the age level of the operator? Are the effects of color display usage related to the experience level of the operator? Are the effects of color display usage related to the time of day of data entry? These research questions were considered by separate analysis of phase 2 and phase 3 data. The analyses are discussed simultaneously for both phases. Initially the required ANOVA technique assumptions were checked and the findings discussed. The research questions are then resolved using the appropriate models.

ANOVA Assumptions

The underlying criterion of independence, homogeneity of variances, normality and additivity of the ANOVA technique were examined for aptness of its application to the models: TERMGP x AGEGP x TOD vs CSESTIME and TERMGP x EXPLV x TOD vs CSESTIME. Each of these assumptions are discussed.

Independence. This required assumption was met due to several factors inherent in the experimental design. The operators worked separately on their own unique set of data entry forms. Also, data were collected only on new applicants and once the initial entry was made by an operator, this applicant was no longer considered new.

Hence two data points could not be collected on the same application. Therefore it is assumed for this research that the data were independent.

Homogeneity of Variances. Compliance with the equality of variances assumption was investigated using Levene's L test for the 3-factor models under scrutiny. For both models this test indicated a conclusion of rejection of the null hypothesis that the variances were equal. Therefore the Welch statistic was considered for use to answer the research questions. However the Welch technique available considered only two factors. Therefore, the models were reduced to 2-factor models. This allowed application of the Welch technique. The 2-factor models considered were TERMGP x AGEGP, TERMGP x EXPLV, and TERMGP x TOD. The results of Levene's L test for these new models are presented in Table 6.9, Homogeneity of Variances for the CSESTIME Variable. This table is constructed and interpreted similarly to Table 6.7 discussed in the previous ANOVA Assumption section for the Error Count analysis. For the three models using phase 2 data, rejection of the null hypothesis that the variances were equal was indicated with $p<.05$. This was also the case for two of the models TERMGP x AGEGP and TERMGP x TOD, using phase 3 data. For these the Welch

Table 6.9

HOMOGENEITY of VARIANCES
for the
CSESTIME VARIABLE

Phase 2 (P2) and Phase 3 (P3)

Ho: Cell Variances Are Equal

Reject if $p \leq .05$

TERMGP x AGEGP p Value (P2/P3)	TERMGP x EXPLV p Value (P2/P3)	TERMGP x TOD p Value (P2/P3)
.01*/.02*	.00*/.28	.01*/.02*

* Rejected at assumed significance level of .05

statistic was planned for analysis as just described. For the TERMGP x EXPLV model using phase 3 data the p value of .28 indicated considerable risk in assuming that the variances were not equal. For this model the Welch statistic was not considered.

Normality. Normality of the residuals in each cell of each 2-factor ANOVA model was investigated using the Kolmogorov D-Statistic procedure in SAS. Assuming a significance level of .05, the statistic indicated rejection of the null hypothesis that the distribution of the residuals was normal. The coefficients of skewness and kurtosis were calculated to investigate if the departure from normality was of extreme form. Table 6.10, Departure from Normality for the CSESTIME Variable, lists the coefficients of skewness and kurtosis. The table lists these values for each of the three models with phase 2 and phase 3 values separated by a slash (/). For all models, the coefficient of skewness was less than 1.4 and the coefficient of kurtosis was less than 2.5. Therefore it was assumed the departure from normality was not of extreme form and the normality assumption satisfied.

Additivity. Additivity was the final criterion investigated. This property was tested by analyzing the interaction terms in each model (Montgomery, 1976; Neter

Table 6.10

DEPARTURE from NORMALITY for the CSESTIME VARIABLE

Phase 2 (P2) and Phase 3 (P3)

TERMGP x AGEGP		TERMGP x EXPLV		TERMGP x TOD	
Skewness (P2/P3)	Kurtosis (P2/P3)	Skewness (P2/P3)	Kurtosis (P2/P3)	Skewness (P2/P3)	Kurtosis (P2/P3)
1.3/1.2	1.9/1.5	1.3/1.2	1.8/1.6	1.2/1.2	1.8/1.5

and Wasserman, 1974). The actual ANOVA table is discussed below in Results and Conclusions.

For those models whose cells satisfied the homogeneity of variance criterion, the interaction term was tested directly for significance at the .05 significance level. This method was only applicable for the TERMGP x EXPLV model using phase 3 data. At this level, the interaction effect was found to be statistically significant. Hence the model was assumed not to be additive. For this model, the main effects' statistics have little practical meaning. The literature suggests holding *one factor* constant while applying the ANOVA technique to the other in order to draw conclusions (Montgomery, 1976; Neter and Wasserman, 1974). This was followed in accomplishing analysis of the TERMGP x EXPLV model.

For those models whose cell's data supported unequal variances and thus the Welch technique planned for main effects analysis, the ANOVA interaction statistics were analyzed but with a more stringent criterion. To compensate for a possible fluctuation in the F statistics for the interaction factor in these models, the F statistics were increased by 17 per cent and the p value recalculated. The level of significance was considered at

.05. This method was required for all three models using phase 2 data and two of the phase 3 models: TERMGP x AGEGP and TERMGP x TOD. All five of these models' recalculated p values indicated considerable risk in assuming the interaction terms were significant. Therefore it was assumed that the main factors of these models were additive.

Conclusions. Independence was applicable for all models considered for both phase 2 and phase 3 data. The 3-factor models failed Levene's L test for homogeneity of variances, hence the Welch technique was planned for main effects analysis. As the available Welch technique considered only 2-factor models, the original models were redefined. After investigating the normality of residuals for the CSESTIME variable for the 2-factor models, deviation from normality was found not to be of extreme form. All models except TERMGP x EXPLV using phase 3 data were found to be additive. For this model, main effects analysis was accomplished at each level of the EXPLV term. For the other models, the statistics derived from the ANOVA and/or Welch techniques were employed for analysis. The results of the analysis are presented next for phase 2 and phase 3 data.

Results and Conclusions

The ANOVA and/or Welch technique were applied to the appropriate models to investigate the effect color had on the time required for data entry. The models considered were TERMGP x AGEGP, TERMGP x EXPLV and TERMGP x TOD vs CSESTIME. The analytical findings are detailed in Table 6.11, ANOVA and Welch Results. The table is similar to Table 6.8 and lists the calculated values for each model for the ANOVA and/or Welch techniques. If the p value was less than or equal to .05, the null hypothesis of nonsignificance of the factor was rejected. The results and conclusions are discussed for each of the research questions.

Color vs Session Time. Does color display terminal usage affect operator time to accomplish the data entry task? The main factor TERMGP was analyzed for all three models and each phase of data. The levels of TERMGP were color and monochrome display terminal. The six p values for the ANOVA technique are presented in Table 6.11. The only value that indicated statistical significance was for the TERMGP x AGEGP model using phase 2 data with p=.02 for the ANOVA technique and p=.00 for the Welch technique. These results imply that when removing the variance due to operator age, the terminal on which data was entered

Table 6.11

ANOVA and WELCH RESULTS

CSESTIME DEPENDENT VARIABLE

Phase 2 (P2) and Phase 3 (P3)

Ho: Factor equal zero

Reject if p≤ .05

Values / Model Factors	Mean SS (P2/P3)	df (P2/P3)	F or W (P2/P3)	p (P2/P3)
TERMGP & AGEGP				
ANOVA				
TERMGP	6952.57/ 724.39	1/1	5.18/0.44	.02*/.51
INTERACTION	17.41/ 153.59	1/1	0.01/0.09	.91 /.76
ERROR	1340.92/1643.90	2181/1807		
WELCH				
MAIN EFFECTS		3,732/3,702	7.17/0.96	.00*/.41
TERMGP & EXPLV				
ANOVA				
TERMGP	1241.68/ 386.51	1/1	0.91/0.24	.34 /.63
INTERACTION	1322.03/14763.2	1/1	0.97/9.03	.32 /.01*
ERROR	1359.87/1634.33	2181/1807		
WELCH				
MAIN EFFECTS		3,959/N/A	0.91/N/A	.44 /N/A
TERMGP & TOD				
ANOVA				
TERMGP	2894.84/ 553.79	1/1	2.13/0.34	.14 /.56
INTERACTION	956.37/4768.69	1/1	0.70/2.90	.40 /.09
ERROR	1358.34/1643.57	2181/1807		
WELCH				
MAIN EFFECTS		3,1173/3,999	1.98/1.09	.12 /.35

* Rejected at assumed significance level of .05

significantly affected the time needed to accomplish the data entry task. This finding was not supported by the phase 3 data where p=.51 indicated no evidence of significant difference between the TERMGP levels. To interpret this result for practical significance, the means of the two TERMGP levels were compared for each phase of data. For phase 2 data the mean time for task completion using the monochrome display terminal was 1.7 per cent (2 seconds) less than the mean time using color. For phase 3 data, the mean time for task completion using the color display terminal was 0.9 per cent (1 second) less than the mean time using monochrome. The data are inconsistent as to which display terminal usage requires less time to complete the data entry task. The data do in some cases support a significant affect on session time by TERMGP but for the operators and the data entry task tested, the difference is minimal.

<u>Color x Age vs Session Time</u>. Is the effect of color display terminal usage on operator performance as measured by time significantly different for particular levels of operator age? The interaction factor in the TERMGP x AGEGP model for phase 2 and phase 3 data was used to investigate this question. The levels of TERMGP were color and monochrome. The levels of AGEGP were the

younger group of operators (35 years of age or less) and the older group of operators (greater than 35 years of age). The results of testing the null hypothesis that the interaction factor means are equal are presented in Table 6.11. The p value of .91 using phase 2 data indicated considerable risk in assuming that the means are significantly different. The p value using phase 3 data of .76 strongly supports this result. Hence, the data contains insufficient evidence to conclude that color display usage significantly changes operator session time for either the younger or the older level of operators.

Color x Experience vs Session Time. Is the effect of color display terminal usage on operator performance as measured by time significantly different for particular levels of operator experience? The interaction factors in the TERMGP x EXPLV model for phase 2 and phase 3 data were used to investigate this question. The levels of TERMGP were color and monochrome. The levels of EXPLV were less experienced operators (2 years or less experience) and the more experienced operators (more than 2 years experience). Experience referred to operator on the job experience accomplishing the specific data entry task used in this research. The results are presented in Table 6.11. Phase 2 data with p=.32 did not indicate that the means for this

factor were significantly different. However the p value of .01 using phase 3 data did imply additivity could not be assumed. To further investigate this model, analyses was accomplished by holding the EXPLV factor at a constant level and applying the ANOVA technique to the new model. For the less experienced operators, p=.03 indicated rejection of the null hypothesis that the TERMGP mean were equal. The means show a 6.3 per cent (7 second) advantage for less experienced operators when they used the monochrome display terminal. For the more experienced operators, the TERMGP levels also indicated a significant difference in performance with p=.02. The more experienced operators decreased their session time by 4.7 per cent (5 seconds) when using the color terminal. Phase 3 data did support that the effect of color display usage on operator performance as measured by time was significantly different for both the less and more experienced levels of operators. Although statistically significant, the differences were small. These data supported an advantage for the less experience operators using the monochrome display terminal and an advantage for the more experienced operators using the color display terminal.

<u>Color x Time of Day vs Session Time.</u> Is the effect

of color display terminal usage on operator performance as measured by time significantly different for particular levels of time of day of data entry? The interaction factors of the TERMGP x TOD model for phase 2 and phase 3 data were used to investigate this question. The levels of TERMGP were color and monochrome. The levels of TOD were morning (entries prior to noon) and afternoon (entries at or after noon). The results are presented in Table 6.11. Both for phase 2 ($p=.4$) and phase 3 ($p=.09$) data, considerable risk is indicated in assuming that the means are not equal. Hence, the data contains insufficient evidence to conclude that color display usage significantly changes operator session time for either morning or afternoon entries.

Conclusions

This chapter addressed the hypothesis that color display terminal usage to accomplish a data entry task affects the performance of an operator experienced with that task. Extensive analyses were accomplished and discussed using two phases of data to investigate this hypothesis. The analyses involved three models: TERMGP x AGEGP, TERMGP x EXPLV and TERMGP x TOD. These models were investigated with respect to the dependent variables of error count and session time. The form of the error

variable analyzed was the error count ratio rank transformation (RCNTRAT). This variable did not allow analysis of TERMGP x TOD. The form of the time variable analyzed was the corrected session time (CSESTIME). The ANOVA and/or Welch techniques were justified and used as appropriate to accomplish the analyses. The statistically significant results supported by the data are consolidated in Table 6.12, Final Results of Analysis of Color. The p values are given in this table for all models and factors considered. All statistically significant results are highlighted with an asterisk (*).

For the error variable, RCNTRAT, no statistically significant results were indicated. The data supports the conclusion that the use of a color display terminal by operators experienced with the data entry task tested does not significantly affect the number of errors committed.

For the session time variable, CSESTIME, some results indicated statistical significance. This was true for the main factor of TERMGP in the TERMGP x AGEGP model using phase 2 (P2) data. Monochrome display terminal usage was found to be 1.7 per cent (2 seconds) faster than color. Also statistical significance was indicated for the interaction factor in the TERMGP x EXPLV model using phase 3 (P3) data. Holding EXPLV levels constant, it was found

Table 6.12

FINAL RESULTS of ANALYSIS of COLOR

p Values of Models

Phase 2 (P2) and Phase 3 (P3)

H0: Means Are Equal

Reject if $p \leq .05$

Variable / Factors	ERROR RATE P2	ERROR RATE P3	SESSION TIME P2	SESSION TIME P3
TERMGP & AGEGP				
TERMGP	.07	.31	.02*	.51
TERMGP & EXPLV				
TERMGP	.19	.82	.34	I
≤ 2 years	-	-	-	.03*
> 2 years	-	-	-	.02*
TERMGP & TOD				
TERMGP	-	-	.14	.56

* Rejected at assumed significance level of .05
I Interaction significant, hold levels constant
\- p values not applicable for this factor of the model

that for the less experienced group of operators (2 years or less experience) monochrome display terminal usage was 6.3 per cent (7 seconds) faster than color. For the more experienced group (more than 2 years experience) color display terminal usage was 4.7 per cent (5 seconds) faster than monochrome. These results are not supported by the other phase of data in both cases just discussed.

The conclusion supported by these findings was that even for the statistically significant results, the difference in time to accomplish the task for the two terminal types was minimal. In some instances monochrome terminal usage was faster and in others, color was faster. However, in all cases the difference was never larger than 6.3 per cent (7 seconds).

VII. SPECIAL EXTENSIONS OF THE EXPERIMENT

Introduction

During the final three weeks of the experiment some special extensions of the research were accomplished. There were two purposes underlying these extensions. One purpose was to validate some areas of concern in the experimental design. These areas included the possible influence on operator performance of physical differences between the terminals used in this study other than the presence of color. Although the physical differences in size, keyboard, etc. were minor, it was desired to investigate the possible affects on operator performance. Another area was that of the possible existence of a learning curve in the data and/or change in the effects on operator performance of the independent variables over time. The second purpose of these special extensions was to emulate as closely as possible the previously cited consultant study in an attempt to verify some of their stated results. Each of these extensions is explained in detail, the analysis approach discussed and the results and conclusions presented.

Extended Daily Terminal Use

Introduction

The first extension was designed to address two research questions. Do the physical differences of the terminals used in the current research other than the presence of color significantly affect operator performance as measured by session time? Does the amount of time spent continuously entering data via a computer terminal significantly affect operator performance as measured by session time? This extension consisted of one operator interfacing with a computer terminal continuously for two hours, part of which involved the new applicant entry task. The time requirement was similar to that of a consultant study. To insure a continuous two hours, this operator was relieved of all other office duties. This operator performed the new applicant entry task at three intervals within the two hour period during which data were collected. These intervals occurred at the beginning, middle and end of each block of time. These intervals became the levels for the first independent variable under consideration called productivity group (PRODGP). This operator worked using one of three terminal configurations for a period of one week each. During the first week, the task was accomplished on a

monochrome terminal and 59 data points were collected. The color terminal was used during the second week, with 57 data points collected. The final week's configuration was the same color terminal as the second except with the color switch in the off position. Sixty one data points were collected. This latter configuration actually presented two colors, green and white, rather than the usual four: green, blue, red and white. Green appeared during data entry and any errors detected by the computer program were presented in white. These three configurations were the levels of the independent variable of terminal *configuration* (CONFIGR). The dependent variable considered was corrected session time (CSESTIME). Like prior analyses, these data were analyzed for aptness of using the ANOVA technique. These are discussed followed by the results and conclusions derived from analyzing the data for evidence of affect on performance.

ANOVA Assumptions

The four assumptions of independence, normality, homogeneity of variances, and additivity required for application of the ANOVA technique were investigated for the model PRODGP x CONFIGR vs CSESTIME.

Independence. The method of data collection was identical to that of the preceding phases. Therefore the

data collected for this extension were also assumed to be independent.

Normality. Normality of the residuals in each cell of the model was examined using the Kolmogorov D-statistic. Assuming a significance level of .05, the statistic indicated rejection of the null hypothesis that the distribution of the residuals was normal. The coefficients of skewness and kurtosis were calculated for determination if the departure from normality was of extreme form. These coefficients are listed in Table 7.1, Extended Daily Terminal Usage Departure from Normality. The first line of this table gives the coefficients with respect to the CSESTIME dependent variable. Since the magnitude of the coefficient of skewness was not less than 1.4 and the coefficient of kurtosis was not less than 2.5, the departure from normality was assumed of extreme form. A rank transformation (Conover and Iman, 1976) was considered to allow the data to more closely satisfy this assumption. The normality test was rerun for this new dependent variable of ranked corrected session time (RCTIME). It still indicated rejection of the null hypothesis at p=.01. However the coefficient of skewness and kurtosis (second line of values in Table 7.1) now indicated that the distribution was sufficiently

Table 7.1

EXTENDED DAILY TERMINAL USAGE

DEPARTURE from NORMALITY
for the
TIME VARIABLE

Model / Variable	PRODGP x CONFIGR	
	Skewness	Kurtosis
CSESTIME	2.50	9.30
RCTIME	- .04	1.10

symmetrical and bell-shaped to satisfy the normality assumption. Therefore, for the remaining analysis the transformed variable (RCTIME) was considered the dependent variable.

<u>Homogeneity of Variances.</u> The assumption of homogeneity of variances between each cell of the ANOVA model was verified using Levene's L test for equality of variances. The result indicated failure to reject the null hypothesis of equal variances with $p=.17$. The assumption was made that the variances were equal.

<u>Additivity.</u> This property was tested by analyzing the interaction terms in the model (Montgomery, 1976; Neter and Wasserman, 1974). Since the homogeneity of variances assumption was satisfied, the interaction term was tested directly for significance at the .05 level. The p value of .74 indicated considerable risk in assuming that the interaction term was statistically significant. Therefore the model was assumed additive and the ANOVA technique was applied to the model PRODGP x CONFIGR vs RCTIME.

<u>Results and Conclusions</u>

The ANOVA technique was applied to the model to gain insight into the research questions. The analytical findings are presented in Table 7.2, Extended Daily

Table 7.2

EXTENDED DAILY TERMINAL USAGE

ANOVA RESULTS

RCTIME DEPENDENT VARIABLE

Ho: Factor equal zero

Reject if $p \leq .05$

Values / Model Factors	Mean SS	df	F	p
TERMGP & AGEGP				
ANOVA				
PRODGP	4063.62	2	1.52	.22
CONFIGR	563.68	2	0.21	.81
INTERACTION	1306.55	4	0.49	.74
ERROR	2665.12	168		

*Rejected at assumed significance level of .05

Terminal Usage ANOVA Results. The table lists the calculated values for the ANOVA technique. The results are discussed for each research question.

Terminal Physical Differences. Do the physical differences of the terminal used in the current research other than the presence of color significantly affect operator performance as measured by session time? The main effects factor CONFIGR was considered to address this question. The resulting p=.81 implied considerable risk in assuming that the mean session time for each terminal configuration was not equal. The data do not support that any terminal physical differences are significantly affecting operator performance as measured by session time.

Continuous Terminal Usage. Does the amount of time spent continuously entering data via a computer terminal significantly affect operator performance as measured by session time? PRODGP was the factor analyzed to address this question. With p=.22 considerable risk was indicated in assuming that there was a significant difference in operator performance level during a two hour period of continuous terminal usage. Thus it is implied that the findings of the research are applicable to situations that require the operator to work at the terminal for a period

of time as long as 2 hours.

Color Switch OFF vs Monochrome

Introduction

The second extension was driven by a comment in a consultant study. This study used the color terminal with the color switch off to simulate monochrome. The comment made was

> that the productivity differences between color and monochrome may be understated as a result of the "white" error messages providing a "color" advantage when processing in monochrome (Shafer, 1982, p. 10).

There were two research questions addressed. Does color display terminal usage with the color switch off have a significantly different effect on operator performance as measured by session time from that of monochrome display terminal usage? Do the physical differences of the terminals used in the current research other than the presence of color significantly affect operator performance as measured by session time. These research questions were addressed using four operator's data collected over a period of two weeks. During the first week 139 data points were collected. Two of the operators accomplished the data entry task using a monochrome computer terminal while the other two operators used the color computer terminal with the color switch in the off

position. During the second week the operators switched terminals and 93 data points were collected. The two terminal configurations comprised the two levels of the TERMGP independent variable in the model. The AGEGP and TOD variables were also supported by this arrangement of operators. The ANOVA model considered for analysis was TERMGP x AGEGP x TOD vs CSESTIME. A discussion of the ANOVA technique assumptions with respect to this model is presented first, followed by the analytical results and derived conclusions.

ANOVA Assumptions

Independence. The data were assumed to be independent due to the manner of collection as previously discussed.

Normality. The Kolmogorov D-statistic was used to test the null hypothesis that the residuals in the cells of the model were distributed normally. These results are presented in Table 7.3, Color Switch Off vs Monochrome Departure from Normality, below the p Value header for week 1 and week 2 of data collection. For week 1, p=.01 indicated rejection of the null hypothesis. Hence the coefficients of skewness and kurtosis were calculated. These are presented in the last two columns of Table 7.3. Since the coefficient of skewness was less than 1.4 and

Table 7.3

COLOR SWITCH OFF vs MONOCHROME

DEPARTURE FROM NORMALITY
for the
CSESTIME VARIABLE

Model / Week	TERMGP x AGEGP x TOD		
	p value	Skewness	Kurtosis
1	.01*	.73	.23
2	.15	N/A	N/A

* Rejected at assumed significance level of .05

the coefficient of kurtosis was less then 2.5, it was assumed that the departure from normality was not of extreme form and the normality assumption satisfied. The second week of data resulted in a p value of .15 for the testing of the null hypothesis that the distribution of the residuals was normal. This implied considerable risk in assuming that the distribution was not normal. Hence the normality criterion was assumed satisfied for this model.

Homogeneity of Variances. Homogeneity of variances of the cells of the ANOVA models was verified using Levene's L test for equality of variances. For both weeks of data considerable risk was associated with concluding that the variances were not equal at p=.14 and p=.28. The assumption was made that the variances were equal.

Additivity. This property was investigated by analyzing the interaction terms in the model. The results are presented for both weeks as the last three lines of values in Table 7.4, Color Switch Off vs Monochrome ANOVA Results. Since the homogeneity of variances was satisfied for both weeks, the interaction terms were tested directly for significance at the .05 level. As all six p values were greater than .05, considerable risk was indicated in assuming that any of the interaction terms were

Table 7.4

COLOR SWITCH OFF vs MONOCHROME

ANOVA RESULTS
for the
CSESTIME DEPENDENT VARIABLE

Week 1 (W1) and Week 2 (W2)

Ho: Factor equal zero

Reject if $p \leq .05$

Values / Model Factors	Mean SS (W1/W2)	df (W1/W2)	F (W1/W2)	p (W1/W2)
TERMGP&AGEGP&TOD				
ANOVA				
TERMGP	85.27/4142.57	1/1	0.06/2.93	.81/.09
TERMGPxAGEGP	524.62/5463.24	1/1	0.37/3.87	.55/.06
TERMGPxTOD	5548.69/4812.46	1/1	3.88/3.41	.06/.07
TERMGPxAGEGPxTOD	816.71/ 279.92	1/1	0.57/0.20	.45/.66
ERROR	1430.06/1411.69	131/85		

*Rejected at assumed significance level of .05

significant. Therefore the model was assumed additive with respect to both weeks of data and the ANOVA technique was applied to analyze the TERMGP x AGEGP x TOD vs CSESTIME model.

Results and Conclusions

The ANOVA technique was applied to the model to resolve the research questions. The analytical findings are presented in Table 7.4 and are discussed for each research question.

Color Switch Off/Physical Terminal Differences. Does color display terminal usage with the color switch off have a significantly different effect on operator performance as measured by session time from that of monochrome display terminal usage? Do the physical differences of the terminals used in the current research other than the presence of color significantly affect operator performance as measured by session time? The main effects factor TERMGP derived from each of the two week's data was analyzed to address these research questions. The levels of TERMGP considered were color with the color switch off and monochrome. The p values of .81 and .07 for week 1 and week 2 respectively indicated considerable risk in assuming that the TERMGP factor was significant. Hence the data provide insufficient evidence

of a difference in operator performance between using the color display with the color switch off and the monochrome display.

Randomness of Error Rate and Session Time Over Time

The final extension of the research involved the analysis of randomness of the data over time for individual operators to investigate the possible existence of a learning curve in the data and/or change in the effects on operator performance as measured by either error count (COUNT) or session time (CSESTIME) over an extended period of time. Four operators were asked to continue using the terminals they used during the previous five weeks for another three weeks. This allowed for data collection over a period of eight weeks. Operators 3 and 8 continued using the color terminals. Five hundred eleven data points were collected. Operators 4 and 5 continued using the monochrome terminals, entering a total of 500 new applicants. The phenomena were analyzed for each operator using regression analysis techniques. The data were plotted allowing for visual inspection of the possible relationship, the slope in the linear regression model estimated and tested for significance, and the second order coefficient estimated in the quadratic model and tested for significance (Lewis and Ford, 1983).

The plots of COUNT vs DATE and CSESTIME vs DATE showed no obvious trends for any of the four operators. These plots are presented as Figure 7.1, COUNT vs DATE Linear Model, Figure 7.2, COUNT vs DATE Quadratic Model, Figure 7.3, CSESTIME vs DATE Linear Model, and Figure 7.4, CSESTIME vs DATE Quadratic Model. In each of these figures, the daily averages are shown for each of the four operators with the data points represented by the operator number. The equations are sketched for each operator. The slope and second order coefficients were estimated and their significance checked with a t-test using the linear and *polynomial regression procedures*. The null hypothesis tested was that these model coefficients equaled zero. The results of these tests are presented in Table 7.5, Data Randomness of COUNT vs DATE, and Table 7.6, Data Randomness of CSESTIME vs DATE. The coefficient for the slope and the p value for the significance test of the slope in the linear model are presented in the first two columns of the table. The final columns delineate the second order coefficient and the p value for testing the significance of the second order coefficient in the quadratic model. As the values of p are all greater than .05, the assumed significance level, considerable risk is implied in assuming that either COUNT vs DATE or CSESTIME

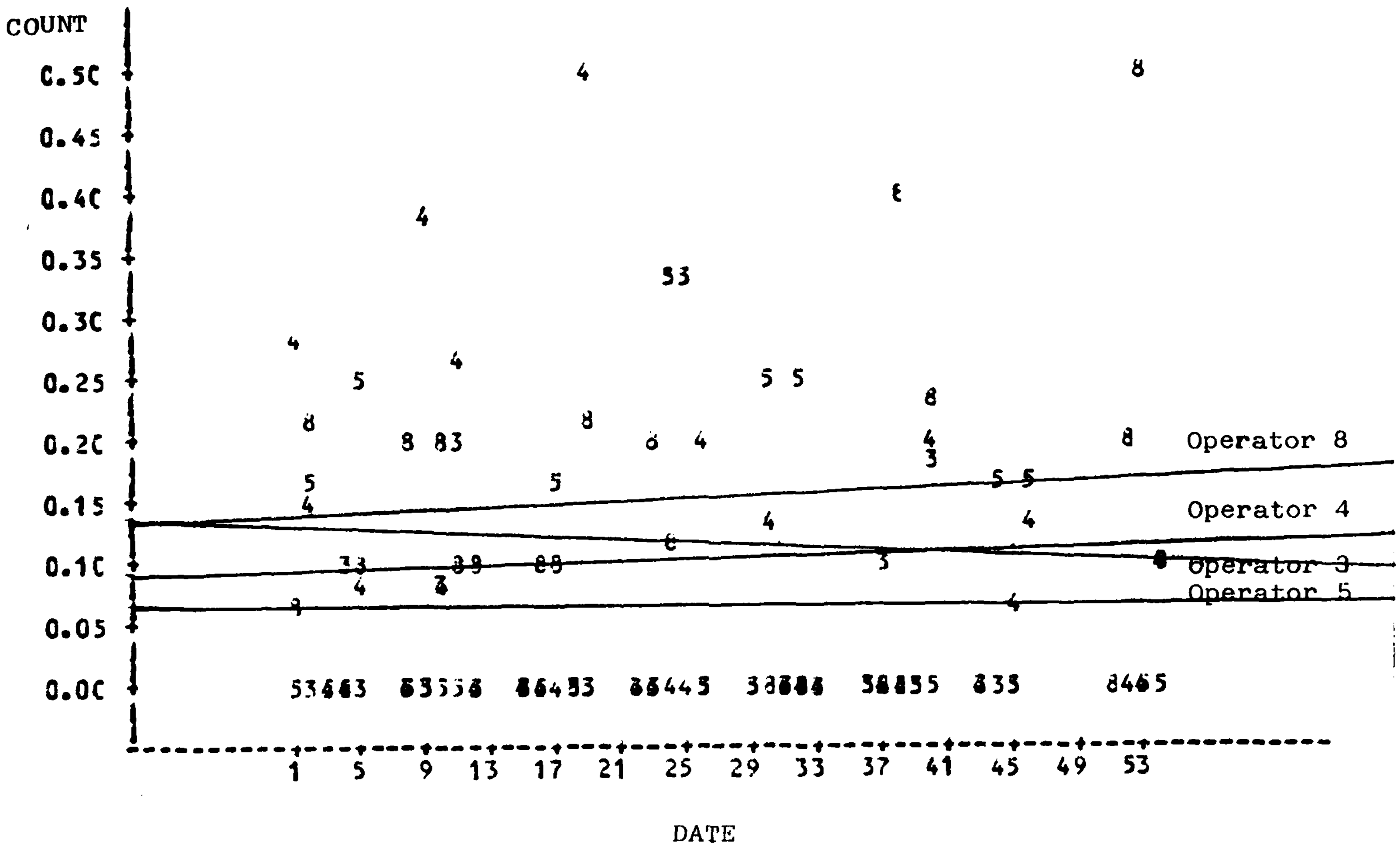


Figure 7.1 COUNT vs DATE Linear Model

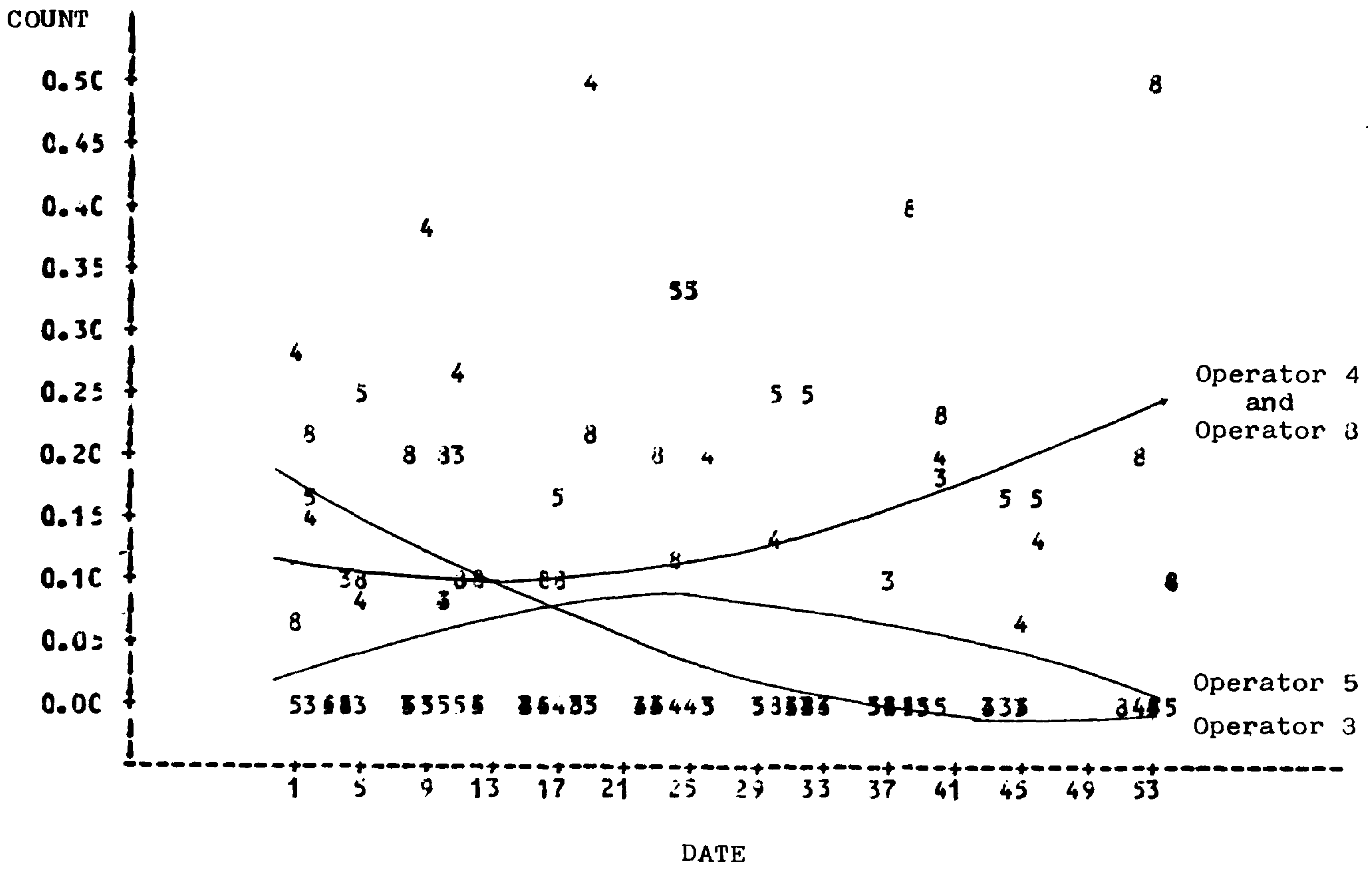


Figure 7.2 COUNT vs DATE Quadratic Model

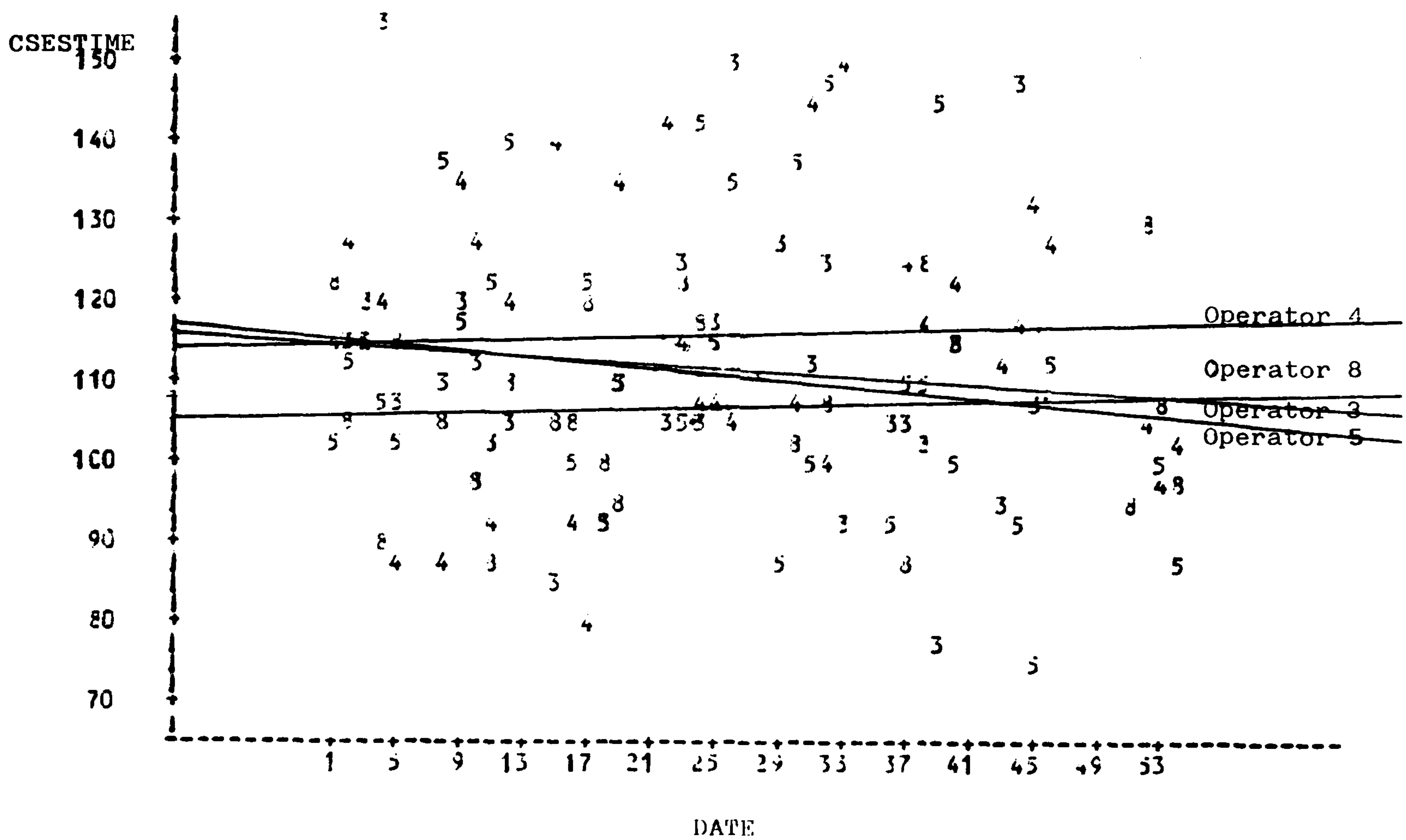


Figure 7.3 CSESTIME vs DATE Linear Model

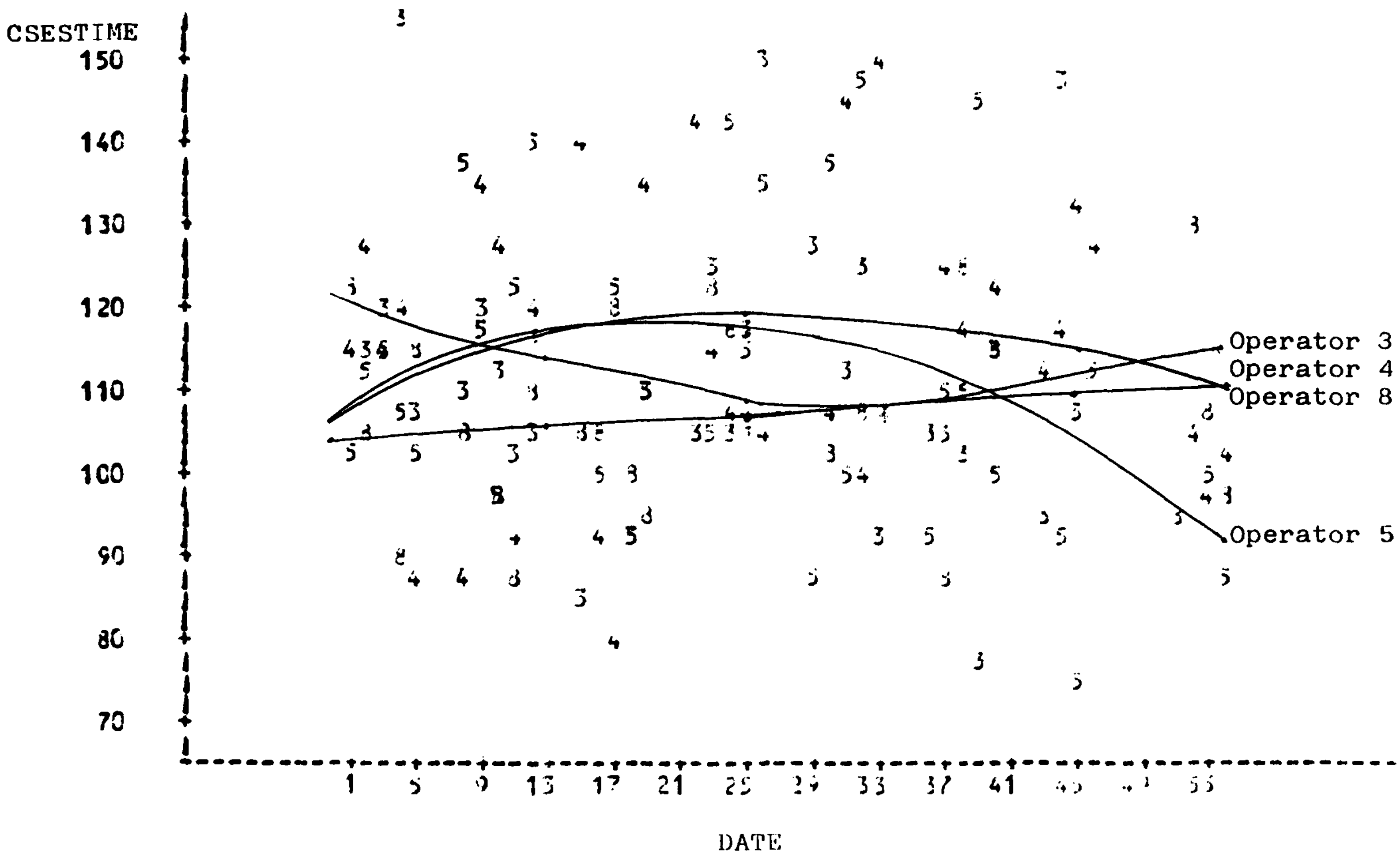


Figure 7.4 CSESTIME vs DATE Quadratic Model

Table 7.5

DATA RANDOMNESS
of
COUNT vs DATE

Ho: Model Coefficient = 0

Reject if $p \leq .05$

Regression Model / Operator Number	First Order Coefficient	First Order p Value	Second Order Coefficient	Second Order p Value
3	-.0031	.18	.0001	.53
4	.0004	.85	.0001	.70
5	.0000	.99	-.0001	.33
8	.0020	.21	.0001	.43

* Rejected at assumed significance level of .05

Table 7.6

DATA RANDOMNESS
of
CSESTIME vs DATE

Ho: Model Coefficient = 0

Reject if $p \leq .05$

Regression Model / Operator Number	First Order Coefficient	First Order p Value	Second Order Coefficient	Second Order p Value
3	-.18	.48	.013	.55
4	.04	.84	-.015	.26
5	-.28	.20	-.025	.08
8	.06	.68	-.001	.90

* Rejected at assumed significance level of .05

vs DATE have evidence of a linear or quadratic relationship. Therefore the data in the case of these operators do not support the existence of either a learning affect or a change in the effects on operator performance over an extended period of time.

Conclusions

Three special extensions of the research were investigated. All three extensions assisted in experimental design validation. The first two extensions aided in further investigation of some stated results in the unpublished consultant study. Conclusions applicable to these two purposes for the extensions conclude this chapter.

The data from the first extension, Extended Daily Terminal Usage, and the second extension, Color Switch Off vs Monochrome, showed no evidence of influence on operator performance due to the physical differences between the terminals used in this study. The data collected in the final extension, Randomness of Error Rate and Session Time Over Time did not support the existence of either a learning affect or a change in the effects on operator performance over an extended period of time.

A consultant study concluded that color display terminal usage improved operator data entry performance

over monochrome. In addition the remark was stated that these improvements may be "understated as a result of the white error messages providing a color advantage when processing in monochrome" (Shafer, 1982, p. 10) To emulate their experimental conditions, the first and second extensions were designed. The first extension data provided no evidence of any statistically significant productivity changes when data was accomplished continuously for a period of 2 hours. The second extension data were collected to compare operator performance using the color display terminal with the color switch in the off position with the monochrome display terminal. These data indicated no support for a statistically significant difference in operator performance when using these terminals. These results are not consistent with those reported in a consultant study, despite the similarities in experimental conditions of terminal type, interface time and data entry task.

VIII. SUBJECTIVE SURVEY ANALYSIS

Introduction

Two surveys, a single terminal evaluation and a multiple terminal comparison, were administered to the operators as a part of this research. The single terminal evaluation survey (Appendix C) given at the end of the first three phases of the study allowed the operator to rate the terminal used during that particular phase. The post study multiple terminal comparison survey (Appendix D) was administered at the end of the seventeen weeks of the study. This survey was to allow the operators to compare the two terminal types. The survey was designed after a survey that was administered in the previously cited consultant study (Appendix E).

The single terminal evaluation survey consisted of fourteen questions. The first nine questions were answered on a five point scale ranging from "strongly disagree" to "strongly agree". Questions 10 and 11 were also answered on a five point scale but this scale ranged from "I love it" to "I hate it". The last three questions were open ended with space provided for the operator's comments. The answers to the first eleven questions on each of the surveys were assigned a scored answer value.

These values ranged from one, if the far left box was the operator's response, to five, if the box to the far right was marked. The values two, three, and four were assigned respectively as the answers were marked higher on the scale from the left. These scored answer values were used to analyze the responses. The comments to the open ended questions were subjectively evaluated.

The multiple terminal post study comparison survey consisted of six statements of interest to the experimenter to which the operators responded "color", "no difference", or "monochrome" as appropriate. The responses to each of these were tallied and compared to those responses from a consultant study. There were also five open ended questions asked as a part of this survey. The comments collected from these are stated.

Terminal Evaluation Survey

Introduction

The survey designed to allow the operators to evaluate a single computer terminal (Appendix C) used in the research was administered at the end of the first three phases of the study. The first phase, lasting four weeks, had all nine of the participants using the existing monochrome display computer terminals for the data entry task. Phase 2, lasting five weeks, assigned an

experimental group, four operators, to the newly installed color display computer terminals while the control group, five operators, continued using the monochrome terminals. Phase 3, lasting five weeks, switched the terminal assignments of phase 2. The analysis accomplished was divided into four categories. The first category used the responses (Appendix H) to the surveys administered after each of the three phases of data collection. This was to check the consistency of the operator's answers, whether the answers for each operator remained the same when they were evaluating the same terminal. This was possible for the monochrome computer terminal as each operator evaluated this terminal twice during the research. The last three categories of analysis used only those responses to the surveys administered following phase 2 and phase 3 data collection. During these phases a portion of the operators were evaluating the color terminal and a portion the monochrome terminal. The first of these three categories of analysis involved the operator's responses to the questions pertaining to terminal similarity. The concern was whether the physical characteristics of the color and monochrome terminals appeared the same to the operators. Then the operator's evaluation of areas possibly relating to the use of a

color and/or a monochrome display terminal was considered. These areas were of primary interest to the research and addressed several questions. Does the type of terminal, color or monochrome, used for data input influence job satisfaction? Does the type of terminal have an effect on how well the operators like the terminal? Does the terminal type influence the effect that interruptions have on the operators when they are performing the data entry task? Is eyestrain and/or headaches a problem for the operators when working on either of the two terminal types? Is physical fatigue a problem for the operators when working on either of the two types of terminals? Finally the open ended questions concerning terminal features were evaluated. Each of these four categories of analysis is discussed separately.

Response Consistency

An analysis of the consistency of the operator's answers to the surveys was accomplished. Consistency was defined as whether or not each operator, when evaluating the same terminal, answered similarly. This was assumed to be the case if the scored answer values given by each operator remained within a difference of one for the two evaluations of the same terminal. The total possible variance of these values was four. This consistency was

investigated for each operator and each question. Analysis was possible only for the monochrome display computer terminal usage as each operator evaluated these terminals twice, either five or ten weeks apart.

The five operators comprising the control group used the computer terminals in the order monochrome, monochrome, color for phases 1 through 3 respectively. The scored answer values for the monochrome computer terminal evaluations following phase 1 and phase 2 were analyzed. These responses remained within a difference of one for each operator and each question.

The four operators referred to as the experimental group used the computer terminals in the order monochrome, color, monochrome for phases 1 through 3 respectively. In this instance the scored answer values for the monochrome computer terminal following phase 1 and phase 3 were analyzed. There was only one exception to the consistency of the answers. The exception arose with operator 2's answer to question 2. The question was, glare on the screen was no problem when you use the terminal. The answer to the phase 1 survey by this operator was a two, implying disagree. The second monochrome terminal evaluation by this operator was a scored answer value of four, implying agree. The verbal comment by this operator

in the interview that followed the phase 3 survey indicated that this difference implied a reaction to use of the monochrome terminal following use of the color terminal, which she used in phase 2 of the research. Her color terminal evaluation for this question was a one, strongly disagree. The strong negative feeling regarding glare for the color terminal thus tended to drive her similar feeling regarding the monochrome toward the positive.

Terminal Similarity

Several of the questions allowed the physical characteristics of the color and monochrome terminals to be evaluated by the operators. The analyses considered only responses to the surveys administered following phase 2 and phase 3 data collection. In addition to considering the difference in the scored answer values for each individual operator, a Wilcoxin signed ranks test was applied. This nonparametric test was used to test the null hypothesis that the operators' mean response for evaluation of the color terminal was equal to their mean response for evaluation of the monochrome terminal. The test applies to paired interval scale measurements taken on the same subject (Daniel, 1978).

Question 1 asked about the work space around the

terminal being adequate. Of concern was to insure that the size of the terminals was similar. None of the operators disagreed that the work space was adequate. Each operator's scored answer value was within a difference of one for each type of terminal. The Wilcoxin test indicated failure to reject the null hypothesis of equal mean responses for evaluation of the two terminals.

Questions 3 and 4 asked about the legibility of the keys and the reach required to access the keys. These questions were to insure that the keyboards were similar, both with respect to key markings and keyboard size. None of the operators disagreed that the keys were easy to read on each of the terminal keyboards. Some of the operators did feel that the keyboards required some strain of their fingers to reach the keys. But these responses were similar regardless of the terminal being evaluated. The Wilcoxin test indicated failure to reject the null hypothesis for each of these questions.

Question 5 was concerned with the computer terminal displays being legible and clear. Again of primary interest was that each operator evaluated both of the terminals similarly in this area. None of the operators disagreed that the color and monochrome terminals were similarly legible and clear. The scored answer values

were within a difference of one and the Wilcoxin test indicated failure to reject the null hypothesis.

Due to the results of these analyses, the experimenter made the assumption that the physical characteristics of the terminals were similar with respect to the areas evaluated by the operators.

<u>Color and Monochrome Terminal Usage Evaluation</u>

The effect of color and/or monochrome computer terminal usage was of concern in six areas relating to operator subjective evaluation. The analysis in each of these areas considered responses from surveys administered following phase 2 and phase 3 data collection. The area of job satisfaction was addressed by questions 9 and 10. Question 11 asked about how well the operator liked the terminal used during the phase of the study being evaluated. The problem of glare was covered by question 2. Question 6 asked for an evaluation of the effect of interruptions when working on the terminals. Operator eyestrain and headaches were addressed with question 7. The area of physical fatigue of the operator when using the two different terminal types was considered by question 8. The results for each of these areas are presented.

Does the type of terminal, color or monochrome, used

for data input influence job satisfaction? Questions 9 and 10 asked the operators if they were satisfied with their jobs and how well they liked their job respectively. None of the scored answer values differed by more than one for each operator's evaluation of the color terminal versus the evaluation of the monochrome terminal. The operators who responded they were not satisfied with their jobs also responded they disliked their job. Similar results were true for operators at the opposite end of the scale. These responses for each operator were the same whether the operator was evaluating the color terminal or the monochrome terminal. The Wilcoxin test indicated failure to reject the null hypothesis of equal mean responses for evaluation of the two terminals for each of these questions. Hence, no evidence was found of an influence of terminal type on job satisfaction.

Does the type of terminal, color or monochrome, have an effect on how well the operators like the terminal? Question 11 allowed the operators to evaluate how well they liked the terminal they used in each phase of the study. For the control group of operators, who went from monochrome terminal in phase 2 to the color terminal in phase 3, none of the operators scored answer values differed by more than one between the terminals. The

experimental group of operators that used the color terminals first and then returned to the monochrome terminals had three operators whose responses were similar for both terminals. The fourth operator responded she disliked the color terminal she evaluated in phase 2 and she liked the monochrome terminal she evaluated in phase 3. This represented a scored answer value difference of two. The Wilcoxin test considering the responses to this question indicated failure to reject the null hypothesis that the mean response was different for the two terminals. Therefore the responses provided insufficient evidence to conclude that terminal type had an affect on how well operators liked the terminal.

How critical is the glare problem for the two types of terminals? An attempt to investigate the problem of glare when the display was in color and when the display was in monochrome was made with question 2. For both phases, the terminals were placed in identical positions. The control group of operators, who used the monochrome terminal in phase 2 and the color terminal in phase 3, evaluation of the glare problem was within a difference of one for each operator's scored answer values. For the experimental group of operators, using the color terminal in phase 2 and the monochrome terminal in phase 3, two of

the operator's scored answer values did not differ by more than one between the two phases. The other two operators strongly agreed that glare was a problem when the data entry was presented on the display in color. They disagreed that glare was a problem when the data entry was presented on the display in monochrome. This represented a scored answer value difference of three. The Wilcoxin test indicated failure to reject the null hypothesis. Therefore the data provided insufficient evidence to conclude that glare was more of a problem for one terminal type than the other, although several operators felt this was the case.

Does the terminal type influence the effect that interruptions have on the operators when they are performing the data entry task? Question 6 addressed this concern. For all but one of the operators the answers were identical for each operator when she evaluated the monochrome display terminal and when she evaluated the color display terminal. The other operator's scored answer value was within a difference of one. Hence, no evidence existed of terminal type influencing the effect that interruptions have on the operators.

Is eyestrain and/or headaches a problem for operators when working on either of the two terminal types? This

was investigated by question 7. For the control group, the scored answer values were all within one between phases for each of the operators. One of the operators in the experimental group evaluated the terminals the same. The other three operators in this group all responded that this was a problem with the color terminal but not with the monochrome terminal. The scored answer value difference for two of these operators was two and for one of the operators was three. However, the Wilcoxin test indicated failure to reject the null hypothesis that the mean responses were equal. Therefore the data do not support that eyestrain and/or headaches were more of a problem with one terminal type than another.

Is physical fatigue a problem for the operators when working on either of the two types of terminals? Question 8 allowed the operators to evaluate this. None of the operators responded that physical fatigue was a problem when using the terminals for data entry. The Wilcoxin test showed no significant response difference between evaluation of the two terminal types.

There were three open ended questions asked as a part of the questionnaire. Two allowed the operator to evaluate what they liked best and least about the terminal. The third asked for any general comments about

the terminal. Few comments were made to any of these questions and those that were made were consolidated. Four of the operators mentioned they liked the color display and the fact that color aided in locating any errors quickly. Three of the operators answered that they disliked the color variations that were used in the display of the four color terminal.

Terminal Comparison Survey

At the completion of the data collection phases of the research a post study survey was administered (Appendix D). This multiple terminal comparison survey consisted of two parts. The first was a series of six statements to which the operators responded the computer terminal they felt was best (color, monochrome or no difference) for that area of concern referred to by the statement. The second part of the survey was five open ended questions concerning the advantages and problems with each terminal type as well as asked for any general comments. The purpose of this survey was to allow the operators to compare the color and monochrome terminals they had used during the previous seventeen week period and to allow the researcher to compare their responses to those from a consultant study. The survey construction was taken from a consultant study. In a consultant study

the survey (Appendix E) was administered to 23 operators who had an average of fifteen months experience using the monochrome terminals. They were allowed to use the color terminals for nine days and then were requested to respond to the questionnaire. The current study administered the survey to nine operators with an average of thirty months experience using the monochrome terminals for the data entry task used in the research. Each of these operators used the color terminals for a minimum of five weeks prior to responding to the survey. For the current study, as discussed in Chapter 3, no reference was made in operator instruction and question periods to color being the variable of concern. Although the color factor could not be hidden, it was never mentioned as being of primary interest. The results from this current study are presented and discussed in comparison/contrast with a consultant study.

The results from the statements portion of the survey for both the current study and a consultant study are presented in Table 8.1, Terminal Comparison Survey Results. The first column of this table lists the statement to which the operator was asked to respond. The responses available to the operator were color, no difference, or monochrome. Percentages for each of these

Table 8.1

TERMINAL COMPARISON SURVEY RESULTS

Percentages of Operators

Statement	Color	No Difference	Mono
I prefer to work with	22.2(83)	22.2(9)	55.5(8)
It was easier to learn to use	0.0(8)	66.6(83)	33.3(9)
I experienced less eyestrain, neckstrain and headaches with	11.1(65)	55.5(26)	33.3(9)
I experienced less fatigue with	0.0(65)	88.8(35)	11.1(0)
I make fewer errors with	33.3(22)	33.3(78)	33.3(0)
I can produce more work with	33.3(27)	33.3(68)	33.3(5)
Totals	17 (45)	50 (50)	33 (5)

categories are presented in the next three columns. In each case the current research value is listed and then the value for a consultant study is presented in parentheses. Total column percentages are also shown.

In response to the first statement, "I prefer to work with", the majority of the operators tested here answered monochrome computer terminal. In a consultant study the majority responded color computer terminal.

The majority of the operators tested in the current study responded that there was no difference in ease of learning the two terminals. This response was the same as that given by the majority of the operators in the earlier study.

The third and fourth statements concerned eyestrain, neckstrain, headaches, and fatigue when using the computer terminals for data entry tasks. The majority of the operators in the current study responded no differences in these effects between terminal types. In a consultant study, the majority of the operators responded that these effects were experienced less when working with the color computer terminal.

The last two statements concerned operator performance. The operators were asked which computer terminal they made "fewer errors with" and which they

could "produce more with". The operators participating in the current study responded equally for all three choices. Three operators felt that color was the best in these areas, three monochrome, and three no difference. The majority of the operators in a consultant study responded no difference to these statements.

The second part of the survey consisted of open ended questions. The responses to these questions are stated. The primary advantage of the color display mentioned by the operators involved in the current study was the ease of finding errors identified by the computer program. The problem area with the color display was felt to be the color contrast, blue and green were used by the terminal as the primary colors for data entry. This was difficult for the operators to get acquainted with and in some cases was felt to cause headaches. The comment made by several of the operators as the primary advantage of the monochrome display terminal was that it was only one color and they felt this was easier on their eyes. No problem areas with the monochrome display terminal were mentioned. Similar comments to these were made on the survey used as a part of a consultant study.

Conclusions

The analysis of the single terminal evaluation survey

administered following the first three phases of the study supported the following results:

1. No evidence that the physical characteristics of the terminals (color and monochrome) were not similar.

2. One operator disliked the color terminals and liked the monochrome terminals.

3. Two operators felt that glare was more of a problem with the color terminal than with the monochrome terminal.

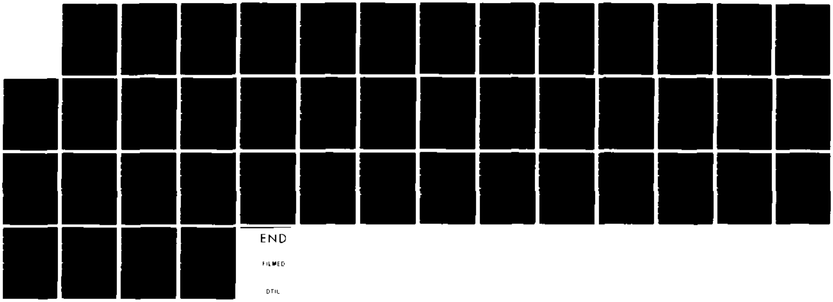
4. Three operators felt eyestrain and headaches were more of a problem with the color terminal than the monochrome terminal.

Although these last three analyses showed operator response being different for the monochrome terminal evaluation and the color terminal evaluation, the Wilcoxin test still failed to reject the null hypothesis that the mean responses were equal for each terminal evaluation. Hence there were some differences in the terminal evaluations, but they were not statistically significant. The primary comments made by both groups of operators to the open ended questions in the survey were that the color display helped locate errors quickly and that the colors used in the color display were not pleasing.

The results of the multiple terminal comparison

AD-A145 568 ANALYSIS OF DATA ENTRY PERFORMANCE: CHROMATIC VERSUS MONOCHROMATIC(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH R L ROSE MAY 84 3/3

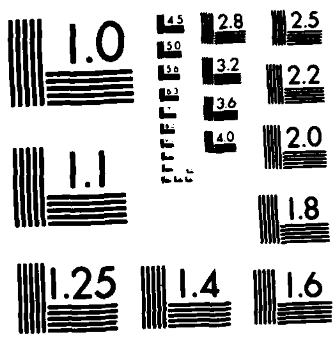
UNCLASSIFIED AFIT/CI/NR-84-48D F/G 5/10 NL



END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

survey administered at the end of the study were discussed and compared with the results from a similar survey used in a consultant study. The majority of the operators in the current study preferred to work with the monochrome display terminals and in the earlier study the color display terminals. The results also showed that the majority of the operators in the current study felt there was no difference between terminals in the amount of eyestrain, neckstrain, headaches, and fatigue experienced. The earlier study showed the majority of the operators felt they experienced less of these symptoms when using the color display terminal. The same responses were found to the open ended questions in this survey as with the terminal evaluation survey. The finding of errors quickly and the disliking of the color combinations with the color display terminal were supported by a consultant study.

In general, the results of the current study were more supportive of the monochrome display computer terminal for data entry than the previous consultant study. Some of the differences in the responses to the survey administered by these two studies can possibly be attributed to several factors. First, the current study attempted to remove, as much as possible, the bias that color was the concern of the research. The operators were

instructed from the initial research session that the research was interested in their total evaluation of the terminals. Secondly, the operators tested in the current research used the color display terminals for a period of five weeks prior to evaluation, as compared to nine days for a consultant study.

IX. CONCLUSIONS AND RECOMMENDATIONS

Introduction

A major relationship of concern to the human factors engineer is that between the design of the machine interface and human productivity. One such interface, gaining wide popularity in business and industry, is the color display computer terminal. Managers have stated that the productivity of experienced operators accomplishing data entry is increased when using a color display versus a monochrome display (Color CRT terminals reduce error rate, Driscoll, 1983; Kelso, 1983; Miller, 1982). However, empirical research substantiating these comments is unavailable and making these draws the conjectures into question.

The research reported here was an attempt to address the benefits of color. Specifically, the effects on efficiency and quality of experienced data entry operators when using color terminals versus monochrome terminals for data entry were studied. The data entry task used in this research was selected as representative of the data entry task accomplished by the users of the color display terminal cited in the literature. The task was currently being accomplished by the Undergraduate Admissions Office,

Arizona State University. It involved the entry of data for applicants requesting admission to the University. Nine operators with a minimum of 1.5 years experience on this job participated. All operators were female ranging in age from 21 to 56 years. The independent variables considered were type of terminal (color and monochrome display), age of operator (35 years or less and greater than 35 years of age), experience level of operator (2 years or less and more than 2 years experience), and time of day of data entry (entries prior to noon and entries at or after noon). The dependent variables were in two categories; objective measures of operator performance and subjective measures of operator attitude. The objective measures were error rate and session time. Operator attitude was measured via two survey instruments. The research data were collected over a period of seventeen weeks. During this time four phases of research were accomplished involving 6688 items of data.

During phase 1, all operators using a monochrome display, a baseline for each operator involved in the study was established. Phase 2 allowed performance to be measured when half the operators used a monochrome display while the other half used a color display. Phase 3 served as verification for any results in phase 2 by reversing

the two groups of operators. Phase 4 experiments were conducted to validate the effects of time on the results and to allow comparison of the current research to an earlier unpublished study by Shafer (1982).

This chapter briefly summarized the findings and states the conclusions from this research for each of the dependent variables: error rate, session time, and operator attitude. Recommendations are then presented for further research considerations.

Summary of Findings and Conclusions

The results and conclusions summarized are based on rigorous analyses using the appropriate statistical methods. These methods included correlation, regression and hypotheses testing through evaluation of the applicable statistics: ANOVA F, Welch W, or Wilcoxin T. The discussion that follows considers the results and conclusions first with respect to error rate, then time, and finally with respect to the subjective measures of operator attitude.

Effects of Color on Error Rate

The effects on operator data entry error rate of the independent variable of terminal type were investigated in conjunction first with the independent variable of operator age and then with operator experience level.

This was accomplished as it was suspected that color terminal usage effects might be age and/or experience level dependent. Phase 2 and phase 3 data were analyzed separately. No statistically significant results were indicated at the .05 significance level as previously shown in Table 6.12. It was concluded that the use of a color display for data entry did not change error rate and that this result holds for all levels of age and experience investigated.

Effects of Color on Session Time

The effects on operator data entry session time of the independent variable of terminal type were investigated in conjunction first with respect to operator age, then with operator experience level, and finally with time of day of data entry. This was accomplished using phase 2 and phase 3 data separately. Some of the results indicated statistical significance at the .05 level of significance as previously shown in Table 6.12. Irrespective of operator age, phase 2 data indicated a small but statistically significant difference in session time means. The mean session time for task completion using the monochrome display was 1.7 per cent (2 seconds) less than the mean session time using the color display. This result was not supported by phase 3 data. The other

result that indicated statistical significance at the .05 significance level was the interaction of terminal type and experience level of the operator. When the levels of experience level were held constant, the mean session time for the less experienced (2 years or less) operators was 6.3 per cent (7 seconds) less when using the monochrome display terminal. The mean session time for the more experienced operators was decreased by 4.7 per cent (5 seconds) when using the color display terminal. These results were supported only by the phase 3 data. The conclusion supported by these findings was that even for the statistically significant results, the difference in time to accomplish the task for the two terminal types was minimal and supported only by one phase of data. In some instances monochrome terminal usage was faster and in others color was faster. But in all cases the difference was never larger than 6.3 per cent (7 seconds).

The effects of color on operator session time were also considered with respect to some findings presented in a consultant study. This study reported positive effects on operator data entry time for the color terminal over monochrome when used for a minimal period of 2 hours. In addition, a consultant study reported that these effects may be understated due to the use of a color terminal with

the color switch off to simulate monochrome. The data from the current study did not support that the color terminal improved operator data entry time even when used for a period of time as long as 2 hours. Also, no support was evident from the current research study data that operator data entry time performance was improved when using a "simulated" monochrome terminal versus a "true" monochrome terminal.

Operator Attitude

Two questionnaires were administered to allow operator attitude to be subjectively measured. The first instrument investigated operator attitude when using one of the terminal types. This instrument considered several operator perceptions: job satisfaction, satisfaction with the terminal, effects of interruptions, glare, eyestrain, headaches, and physical fatigue. The operator perceptions were analyzed and no evidence was indicated of a statistically significant difference in their evaluation of the color versus their evaluation of the monochrome display terminal. The second instrument was a replicate of a survey used in a consultant study. The perceptions were evaluated and compared to that study. In general, the responses from the research reported here were more supportive of the monochrome display terminal than were

the responses from a consultant study.

Conclusions Generalized to Problem Statement

The primary objective of this research was to investigate the effects on performance of experienced operators when using color terminals versus monochrome terminals when accomplishing a data entry task. The task was performed in an environment where the subjects were required to share their attention with other duties of their jobs besides the tested task.

Overall, based on four phases of data collected over a period of seventeen weeks, 6688 data points, the attribute of color in the visual display used to accomplish data entry does not affect error rate or session time of the experienced operator. In addition, the attitude of the experienced operator was more supportive of the monochrome display than the color display.

In general, the results using the objectively measured data are consistent with other empirical studies on the effects of color visual displays. The current research data do not contain sufficient evidence to conclude that color has a unique quality for data entry. These results are not consistent with a consultant study. The discrepancy may be due to several factors. A

consultant study was performed by a productivity consultant specifically for a company that manufactures and promotes color terminals. A consultant study lasted only for a period of 2 weeks and was not accomplished in a workforce environment. These factors may have introduced bias into a consultant study. Subjectively, the research reported here indicated that for experienced data entry operators, the color display is not felt to be an advantage over the monochrome display. These subjective findings are inconsistent with those of previous studies using operators inexperienced with the tested task. The difference is possible attributable to the fact that once an operator becomes experienced with the data entry task using a form presented on the visual display that never changes, the display is seldom referenced.

Areas for Further Research

Recommendations

1. Several of the operators commented that the colors presented by the terminal in this research were not pleasing. Would operator performance be enhanced by use of some other set of colors? What colors? Would the use of more than four colors assist the operator? How many?

2. Ambient lighting is a primary consideration in any environment where visual displays are in use. Does

the type of lighting required differ when using a color display? Is lighting a bigger design issue in the vicinity of color displays due to the lower contrast of colors?

3. The majority of data entry operators currently employed are female. The few research studies accomplished investigating data entry have used female operators. What are the effects of color display usage on the male data entry operator? Are the effects different than for the female data entry operators?

4. Data entry is not always accomplished via a formatted display as was used in the research reported here. Does color have an effect on operator performance when the data entry is performed using other than a formatted display?

5. In the current research, the operator was very familiar with the screen format in which the data was entered. If the format were redesigned, would color aid the experienced operator to learn the new format more quickly than monochrome? If so, would this level of operator performance be maintained?

6. The data entry task discussed in this research was accomplished in conjunction with many other tasks required by the job. What are the effects of color on

performance when the tested task is the only task accomplished throughout the workday?

7. When measuring operator performance, one of the objective measures cited in the literature was error rate. This was one of the dependent variables of the current study. Another aspect of this measure is error type. Are the type of errors made using a color display for data entry different than using a monochrome display? Are the error types committed more or less critical when using the color display?

Conclusions

There continues to be a large gap in the empirical research information concerning the proper use and the effects of the color display terminal for data entry. The reported study attempted to narrow this gap by including Christ's (1975) three key suggestions to be considered in future research concerned with color visual displays: experienced operators, complex task, and real world environmental setting. This research effort and others previously suggested may assist the human factors engineers in their objective to maximize human/machine performance and allow organizations to efficiently use their resources.

BIBLIOGRAPHY

Air University compendium of research topics, 1982-83, Vol. 1. Air University (ATC), Maxwell Air Force Base, Alabama, July 1982.

Altman, J. W. Improvements needed in a central store of human performance data. Human Factors, 1964, 6, 681-686.

Bailey, S. J. Intelligent terminals guide operator to faster, better decisions. Control Engineering, November 1980, 27, 87-92.

Barker, E. & Krebs, M. J. Color effects on human performance: An annotated bibliography (Technical report ONR-CR213-136-1F). Minneapolis, Minn: Honeywell, Inc., (NTIS AD A039318), April 1977.

Barr, A. (editor). SAS user's guide, 1979 edition. North Carolina: SAS Institute Incorporated, 1979.

Belie, R. G. & Rapagnani, N. Color computer graphics - a design and analysis potential scarcely tapped. Astronautics and Aeronautics, June 1981, 19, 41-48.

Berenson, M., Levine, D. & Goldstein, M. Intermediate statistical methods and applications: A computer package approach. New Jersey: Prentice Hall, Inc., 1983.

Blake, M. Time of day effects on performance in a range of tasks. Psychonomic Science, 1967, 9, 349-350.

Box, G., Hunter, W. & Hunter J. Statistics for Experimentation. New York: John Wiley and Sons, Inc., 1978.

Brooks, R. Search time and color coding. Psychonomic Science, 1965, 2, 281-282.

Brown, M. & Forsythe, A. The small sample behavior of some statistics which test the equality of several means. Technometrics, February 1974, 16, 129-132.

Buchanan, M. D. Effective utilization of color in multidimensional data presentations. Advances in Display Technology, 1979, edited by J. R. Parsons, San Diego, Cal. Society of Photo-Optical Instrumentation Engineers.

Bylander, E. G. Electronic displays. New York: McGraw-Hill, 1979.

Cahill, M. & Carter, R.C., Jr. Color code size for searching displays of different density. Human Factors, 1976, 18, 273-280.

Cakir, A., Hart, D. J. & Stewart, T .F. M. Visual display terminals. New York: John Wiley & Sons, 1980.

Calev, F. Enhancing comprehension with color. Mini-Micro Systems, August, 1981, 14, 139-140.

Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Co., 1963.

Carlson, G. Predicting clerical error in an EDP environment. Datamation, 1963, 9, 34-36.

Carter, C. & Cahill, M. C. Regression models of search time for color-coded information displays. Human Factors, June 1979, 21, 293-302.

Carter, R. C. & Carter, E. C. High-contrast sets of color. Applied Optics, August 15, 1982, 21, 2936-2939.

Chapanis, A. Research techniques in human engineering. Baltimore: Johns Hopkins Press, 1959.

Chard, A. Computer display - a clear view of present methods. Process Engineering, September 1979, 116-117.

Chase, W. D. Evaluation of several TV display system configurations for visual simulations of the landing approach. IEEE Transactions on Man-Machine Systems, 1970, MMS-11, 140-149.

Christ. R. E. Four years of color research for visual displays. Proceedings of the Human Factors Society, 21st annual meeting. San Francisco, California, October 1977, 319-321.

Christ, R. E. Review and analysis of color coding research for visual displays. Human Factors, 1975, 17, 542-570.

Christ, R. E. & Corso, G. M. The effects of extended practice on the evaluation of visual display codes. Human Factors, February 1983, 25, 71-84.

Christner, C. A. & Ray, H. W. An evaluation of selected combinations of target and background coding on map-reading performance - experiment V. Human Factors, 1961, 3, 131-146.

Color adds life to computer graphics. Infosystems (USA). January 1981, 28, 24.

Color CRT terminals reduce error rates. Modern Materials Handling. August 6, 1981, 36, 83.

Color tips off stress on finite element display. Machine Design, November 12, 1981, 53, 12.

Colquhoun, W. P. & Corcoran, D. W. S. The effects of time-of-day and social isolation on the relationship between temperament and performance. British Journal of Social & Clinical Psychology, 1964, 3, 226-231.

Conover, W. Practical nonparametric statistics. New York: John Wiley and Sons, Inc., 1971.

Conover, W. & Iman, R. On some alternative procedures using ranks for the analysis of experimental designs. Communications in Statistics-Theory and Methods, A5: 9-15, 1349-1368, 1976.

Cook, T. C. Colour coding - a review of the literature. (NTIS AD A001555), November 1974.

Craig, A. Discrimination, temperature, and time of day. Human Factors, February 1979, 21, 61-68.

Daniel, W. Applied nonparametric statistics. Boston: Houghton Mifflin Company, 1978.

Davies, D. M. & Davies, D. R. The effect of noise and time of day upon differences in performance of two checking tasks. Ergonomics, 1975, 18, 321-336.

Demars, S. A. Human factors considerations for the use of color in display systems (NASA-TM-X-72196, TR-1329). (NTIS N75-14446), January 1975.

Dixon, W. (editor). BMDP statistical software 1981. Los Angeles: University of California Press, 1981.

Dooley, R. P. & Harkins, L. E. Functional and attention getting effects of color on graphic communications. Perceptual and Motor Skills, 1970, 31, 851-854.

Driscoll, M. Director of Admissions, Morristown Memorial Hospital, Morristown, New Jersey. Interview, March 10, 1983.

Durrett, J. & Trezona, J. How to use color displays effectively, April 1982, BYTE 7, 50-53.

Durrett, J., Zwiener,C. & Freund, R. The human factors of color display-based instruction. Proceedings of the National Educational Computing Conference, (ERIC 207 526) 1981, 284-286.

Dvorine, I. Color perception testing charts (second edition). Baltimore, Maryland, 1944.

Friend, D. Color graphics information systems boost productivity. Mini-Micro Systems, May 1980, 13, 181-182.

Granda, R. E. Man/machine design guidelines for the use of screen display terminals. Proceedings of the Human Factors Society, Los Angeles, California, October 1980, 90-92.

Green, B. F. & Anderson, L. K. Color coding in a visual search task. Journal of Experimental Psychology, 1956, 51, 19-24.

Hale, B. Desktop computer clarifies graphics with vivid color. Elect onic Design, May 10, 1980, 28, 161-166.

Hanson, D. C. Introduction to this special issue on the application of color to displays. Proceedings of the Society for Information Display, First quarter 1979, 20, 1-2.

Helwig, J. SAS Introductory guide. North Carolina: SAS Institute Incorporated, 1978.

Helwig, J. T. & Council, K. A. (editors). SAS/graph user's guide, 1981 edition. North Carolina: SAS Institute Incorporated, 1981.

Hitt, W. An evaluation of two different abstract coding methods. Human Factors, July 1961, 3, 120-130.

Hoisman, A. J., Hannah, L. D. & Scharf, E. An experimental investigation of intelligence information display parameters (Report RADC-TDR-63-378). Griffis AFB, New York: Rome Air Development Center, September 1963.

Horsley, J. Colourful computing. Data Processing (G. B.), February 1981, 23(2) 20-23.

How traditional loyalties shift when the office is automated. Management Review, April 1982, 71, 6-7.

Information handling: The major trends. Modern Materials Handling. May 1980, 35, 26-27.

Jacobs, B. A. Humanizing the electronic office. Industry Week, September 21, 1981, 210, 85-86.

Jones, M. R. Color coding. Human Factors, 1962, 4, 355-365.

Judd, D. & Wyszecki, G. Color in business, science and industry. New York: John Wiley & Sons, 1975.

Kanarick, A. & Petersen, R. Redundant color coding and keeping-track performance. Human Factors, April 1971, 13, 183-188.

Karner, C. Perceived versus actual value of color coding. Proceedings, Human Factors Society, 19th Annual Meeting. Santa Monica, California: Human Factors Society, 1975, 227-231.

Keister, R. S. Data entry performance on color versus monochromatic displays. Proceedings, Human Factors Society. Rochester, New York, October 1981, 736-740.

Kelso, B. Data Processing Manager, Wilkens Pipe and Supply Co., Peoria , Illinois. Interview, July 18. 1983.

Kendall, M. & Stuart, A. The advanced theory of statistics. London: Griffin & Company Limited, 1958.

Kochnar, D. S. & Ali, H. M. Age as a factor in combined manual and decision tasks. Human Factors, October 1979, 21, 595-603.

Krebs, M. J. & Wolf, J. D. Design principles for the use of color in displays. Proceedings, Society for Information Display, first quarter 1979, 20, 10-15.

Kunze, K. R. Problems of work in advancing age. Proceedings, Human Factors Society, 19th Annual Meeting. Human Factors Society, 1975.

Levene, H. Robust tests for equality of variances in Contributions to probability and statistics. Olkin, I. (editor). California: Stanford University Press, 1960.

Lewis, B. R. & Ford, R. K. Basic statistics using SAS. New York: West Publishing Company, 1983.

Linksz, A. An Essay on color vision and clinical color-vision tests. New York: Grune and Stratton, 1964.

Mazo, W. S. The facilities planner and the office of the future. Industrial Development, May/June 1982, 151, 36.

McCormick, E. J. & Sonders, M. S. Human factors in engineering and design. New York: McGraw-Hill Book Company, 1952.

McKendrick, J. The office of 1990: Human resources. Management World, January 1982, 11, 14-17.

Mendenhall, W. & Scheaffer, R. Mathematical statistics with applications. North Scituate: Duxbury Press, 1973.

Miller, D. C. Handbook of research design and social measurements. New York: Van Rees Press, 1964.

Miller, F. W. CRT terminals get smarter, cheaper. Infosystems, September, 1981, 28, 100-102.

Miller, F. W. The world of CRT terminals. Infosystems. June 1982, 29, 86-92.

Miller, J. & Rockwell, T. The use of human subjects in human factors research. Human Factors, February 1972, 35-40.

Montgomery, D. Design and analysis of experiments. New York: John Wiley and Sons, Inc., 1976.

Moran, A. S. Jr. Improving office productivity. Credit and Financial Management, October 1981, 83, 31.

Morris, G. Using color in industrial control graphics. Control Engineering, July 1979, 26, 41-45.

Munns, M. Some effects of display symbol variation upon operator performance in aircraft interception. Perceptual and Motor Skills, 1968, 26, 1215-1221.

Murlijacic, T. Computer graphics: A 20 year forecast. Journal of System Management, November 1981, 32, 17-21.

Myers, W. Computer graphics: The need for graphics design. Computer, June 1981, 14, 86-91.

Myers, W. Computer graphics: The need for graphics design. Computer, July 1981, 14, 82-88.

Neary, M. Director of Undergraduate Admissions, Arizona State University, Tempe, Arizona. Interview, February, 1983.

Neter, J. & Wasserman, W. Applied linear statistical models. Illinois: Richard Irwin, Inc., 1974.

Ohlsson, K., Nilsson, L. & Ronnberg, J. Speed and accuracy in scanning as a function of combinations of text and background colors. International Journal of Man-Machine Studied, February 1981, 14, 215-222.

Parsons, H. The scope of human factors in computer-based data processing systems. Human Factors, 1970, 12, 165-175.

Pilla, L. The office of 1990: Automation. Management World, January 1982, 11, 18-20.

Psychology of color. Chemical Business, August 24, 1981, 5.

Ramsey, H. R. & Atwood, E. Human factors in computer systems: A review of the literature. Catalog of Selected Documents in Psychology, May 1980, 10.

Ramsey, H. R., Atwood, M. E. & Kirschbaum, P. J. A critically annotated bibliography of the literature on human factors in computer systems (Technical report SAI-78-070-DEN). Englewood, Colorado: Science Applications, Inc., (NTIS AD A058081), May 1978.

Ray, A. (editor). SAS user's guide: Statistics, 1982 edition. North Carolina: SAS Institute Incorporated, 1982.

Reinhardt, P. (editor). The SAS supplemental library user's guide, 1980 edition. North Carolina: SAS Institute Incorporated, 1980.

Saenz, N. & Riche, C. Shape and color as dimensions of a visual redundant code. Human Factors, 1974, 16, 308-313.

Sanple, R. L. Color your files for productivity. Administrative Management, May 1981, 42, 62-66.

Schatz, W. The latest in executive office gear: A terminal on every desk? Nation's Business, February 1982, 70, 64+.

Schrack, G. F. Current literature in computer graphics and interactive techniques: References. Computer Graphics, February 1980, 13, 381-394.

Schrack, G. F. Current literature in computer graphics and interactive techniques: References, fourth series. Computer Graphics, July 1980, 14, 73-83.

Schrack, G. F. Computer graphics: A key word indexed bibliography for the years 1976, 1977 and 1978. Computer Graphics and Image Processing, September 1980, 14, 24-79.

Schrack, G. F. Computer graphics: A key word indexed bibliography for the year 1979. Computer Graphics and Image Processing, January 1981, 15, 45-78.

Schrack, G. F. Current literature in computer graphics and interactive techniques: References, fifth series. Computer Graphics, March 1981, 14, 200-218.

Schrack, G. F. Computer graphics: A key word indexed bibliography for the year 1980. Computer Graphics and Image Processing, February 1982, 18, 145-187.

Schutz, H. An evaluation of methods for presentation of graphic multiple trends. Human Factors, July 1961, 3, 108-119.

Scott, W. Electronic displays dominating designs. Aviation Week, March 14, 1983, 118, 251-253.

Shafer, J. D. Color versus monochrome display terminals comparative study. Unpublished study conducted for ITT Courier, August 1982.

Shontz, W., Trumm, G., & Williams, L. Color coding for information location. Human Factors, June 1971, 13, 237-246.

Singleton, W. T. Techniques for determining the causes of error. Applied Ergonomics, 1972, 3, 126-131.

Smith, J. A review of literature relating to visual fatigue. Proceedings of The Human Factors Society, Boston, Massachusetts, October 1979, 362-366.

Smith, S. L. Color-coded displays for data processing systems. Electro-Technology, April 1962, 71, 63-69.

Smith, S. L. Color coding and visual search. Journal of Experimental Psychology, 1962, 64, 434-440.

Smith, S. L. Color coding and visual separability in information displays. Journal of Applied Psychology, 1963, 478, 358-364.

Smith, S. L. & Thomas, D. W. Color versus shape coding in information displays. Journal of Applied Psychology, 1964, 48, 137-146.

Smith, S. L., Farquhar, B. B. & Thomas, D. W. Color coding in formatted displays. Journal of Applied Psychology, 1965, 49, 393-398.

Stammerjohn, L. W., Jr., Smith, M. J. & Cohen, G. F. Evaluation of work station design factors in VDT operations. Human Factors, 1981, 23, 401-412.

Stanton, S. Video terminsl under assault. The Arizona Republic, April 12, 1983, A9-A10.

Sublett, J. G. Human component in man-machine systems; a balancing act. Central Engineering, December 1980, 278, 62-64.

Taylor, M. W. Color CRT terminals gain wider popularity. Mini-Micro Systems, December 1979, 12, 86-90.

Teichner, W. H. Color and visual information coding. Proceedings of The Society for Information Display, 1979, 20, 3-9.

Teichner, W. H., Christ, R. E. & Corso, G. M. Color research for visual displays (Report ONR-CR 213-102-4F). Las Cruces, New Mexico: New Mexico State University, Department of Psychology, June 1977.

Teichner, W. H. & Krebs, M. J. Visual search for simple targets. Psychological Bulletin, 1974, 81, 15-28.

Teja, E. Varied color-CRT technologies suit differing graphics applications. Electrical Design News, May 5, 1980, 25, 39-40.

Truckenbroad, J. R. Effective use of color in computer graphics. Comuter Graphics, August 1981, 15, 83-90.

Tullis, T. S. An evaluation of alphanumeric, graphic, and color information displays. Human Factors, October, 1981, 23, 541-550.

Wagner, D. W. Color Coding: An annotated bibliography. Technical Report NWC-TP-5922. China Lake, California: Naval Weapons Center, (NTIS Ad A041061), March 1977.

Waller, L. Roundup: Color monitors blossom. Electronics, April 10, 1980, 53, 153-154.

Waller, R., LeFrere, P. & MacDonald-Ross, N. Do you need that second color? IEEE Transactions on Professional Communication, June 1982, 25, 80-85.

Wedell, J. & Alden, D. G. Color versus numeric coding in a keeping-track task: Performance under varying load conditions. Journal of Applied Psychology, 1973, 57, 154-159.

Whieldon, D. The newest CRTs: Graphics and more. Computer Decision, August 1981, 13, 90-96.

Appendix A
New Applicant Screen Information

```
                 UNDERGRADUATE REGULAR SCREEN          AS OF 05/14/82

AD04      UG ADMISSIONS (RG)              NEXT: ACTION _ ID ________ SCREEN __
   ID: 123-45-6789                   | APP-TYPE: 2 TRANSFER FRESHMAN(1-8 HRS)
 NAME: JONES, JOHN LIONEL, SR        | NO: 1  APP-DATE: 09-24-80  YR/TERM: 81/1
                                     | FEE: Y      TEST: C   ACT:  25   SAT: ___
  ADR: C/O NIELSON                   | RNK/CLS:   20/ 200    10 %
       12623 SOUTHERN AVE            |
  CTY: TEMPE                         | STATUS: A1 UNCONDITIONAL     DT: 11-20-80
   ST: AZ      ZIP: 85282   CTRY: US | COND: A  ADDTN'L DEG        DT:
PHONE:602-946-1998                   | *RDY TO EVAL*      LAST REG S ENRL: 80/1
                                     | ENTRY-LVL: FRESHMAN
  SEX: M  DOB: 12-15-60  ETHNIC: W   | COLLEGE: LA LIBERAL ARTS
 HDCP: N  HLT-PR:  Y VET: N  ATH: N  |  DEGREE: 21 BA
RELIG: 51 NO PREFERENCE INDICATED    | MAJOR: 2205 HISTORY

CITIZEN OF: PC CANARY ISLANDS                   ARIZONA-RES: N
   VISA: IP     PROOF: Y    SPONSOR: ______     TOEFL: 535     FIN-GT: 2
------------------------------------------------------------------------------
  CODE    PREVIOUS INSTITUTION     ST  ATTENDED  TRANSCRIPT  SEM  T/HRS |  GPA
------------------------------------------------------------------------|  2.55
HS 021    ARCADIA SCOTTSDALE       AZ  GRAD 77   C 10-09-80             =====
001083    UNIVERSITY OF ARIZONA    AZ  79-12/80  I 11-15-80  ___   6.0
NOTE: ___________________________  __  ________  _ ________  ___  _____
______________________________________________________________________________
```

Appendix B
New Application Entry Log

NEW APPLICATION ENTRY LOG

Evaluator USERID:__________

Applicant Entry Session				Interruptions (>15 seconds)		
Date	Begin Time	End Time	Terminal #	ASU ID	Hold	Abort

Appendix C
Single Terminal Evaluation Survey

TERMINAL DISPLAY

USER SURVEY

Please answer the following questions to help determine your reaction to the terminal you are currently using. Please place an "X" above the line that best describes your answer to each of the questions.

For Example:

Strongly disagree ___ : ___ : ___ : _X_ : ___ Strongly agree

Thank you.

USERID __________

1. The desk or table space is adequate for documents, listings, etc. needed when you use the terminal.

Strongly disagree ___ : ___ : ___ : ___ : ___ Strongly agree

2. Glare on the screen is no problem when you use the terminal.

Strongly disagree ___ : ___ : ___ : ___ : ___ Strongly agree

3. The terminal keys are easy to read.

Strongly disagree ___ : ___ : ___ : ___ : ___ Strongly agree

4. The terminal keyboard requires no straining of fingers or arms to use.

Strongly disagree ___ : ___ : ___ : ___ : ___ Strongly agree

5. The display on the terminal is legible and clear to read.

Strongly disagree ___ : ___ : ___ : ___ : ___ Strongly agree

6. I found interruptions frustrating when using the terminal.

Strongly disagree ___ : ___ : ___ : ___ : ___ Strongly agree

7. I experienced eyestrain and/or headaches using the terminal.

Strongly disagree ___ : ___ : ___ : ___ : ___ Strongly agree

8. I experienced fatigue using the terminal.

Strongly disagree ___ : ___ : ___ : ___ : ___ Strongly agree

9. I feel satisfied with my job.

Strongly disagree ___ : ___ : ___ : ___ : ___ Strongly agree

10. Which best describes how well you like your job?

I love it ___ : ___ : ___ : ___ : ___ I hate it

11. Which best describes how well you like the terminal?

I love it ___ : ___ : ___ : ___ : ___ I hate it

12. The thing I liked best about the terminal I was using during the past 4 weeks is:

13. The thing I liked least about the terminal I was using during the past 4 weeks is:

14. Other comments I would like to make about the terminal I was using during the past 4 weeks are:

Appendix D
Multiple Terminal Comparison Survey

Please answer the following questions to evaluate the two types of terminals you have been using for the past seventeen weeks (use space on back if needed).

Thank you.

USERID: __________

	Check One		
	Color	No Difference	Monochrome
I prefer to work with	____	____	____
It was easier to learn to use	____	____	____
I experienced less eyestrain, neckstrain, and headaches with	____	____	____
I experienced less fatigue with	____	____	____
I make fewer errors with	____	____	____
I can produce more with	____	____	____

The primary advantage of the color display terminal for me was:

The problem I encountered most often when using the color display terminal was:

The primary advantage of the monochrome (green) display terminal for me was:

The problem I encountered most often when using the monochrome (green) display terminal was:

Other comments I would like to make about the terminals or the research study are:

Appendix E
Courier Study Survey

TERMINAL DISPLAY
USER SURVEY

Please answer the following questions to help determine User Reaction to the Courier Terminals. Do not put your name on the questionnaire. Please return the completed questionnaire to the person who distributed it as soon as possible. Thank you.

I have been working with display terminals for ___YRS., ___MOS., ___WKS.

I have used the color terminal for ___YRS., ___MOS., ___WKS.

	Check One		
	Color	No Difference	Monochrome
I prefer to work with	___	___	___
It was easier to learn to use	___	___	___
I experienced less eyestrain, neckstrain, and headaches with	___	___	___
I experienced less fatigue with	___	___	___
I make fewer errors with	___	___	___
I can produce more with	___	___	___

The primary advantage of the color display for me is:

The problem I encounter most often when using the color display is:

Why do you believe your Company brought in the color terminals:

Appendix F
Operator Consent Form

COLLEGE OF ENGINEERING AND APPLIED SCIENCES
Research Participation Agreement

Type of Research: PhD
Researcher Reynold L. Rose Phone 897-7556
PLEASE READ THE FOLLOWING BEFORE YOU SIGN THE CONSENT FORM
Description of Procedure:

The research will consist of collecting data on your entry of new applicants into the university data base as currently accomplished at this time. Both the existing terminals and newly installed terminals will be utilized. All data will be entered via an assigned terminal or group of terminals using the operators assigned user identification code. It is requested that as a minimum 5 new applications be entered in the morning (8AM - 12PM) and 5 in the afternoon (12PM - 4PM). Entries into a work log provided are requested. Discussion between operators concerning this research should be avoided.

Thank you for your cooperation in this research.

CONSENT

My signature below, in return for the opportunity of participating as a subject in a scientific research investigation, hereby authorizes the performance upon me of the procedure described above. This consent I give voluntarily and after the nature and purpose of the experimental procedure, the known dangers, and the possible risks and complications have been fully explained to me. I knowingly assume the risks involved, and am aware that I may withdraw my consent and discontinue participation at any time without penalty to myself.

Signature:__________________ Date:__________________________

Appendix G
Data Examples

```
83021415:39:04.4015:39:56.40ADUGLL03PADEMD   0516468771
83021415:39:04.4015:40:15.68ADUGLL03PADEMD   0516468771
83021415:39:40.5215:40:40.61ADUGLL07PADDJR   5276960651
83021415:40:16.9515:41:12.56ADUGLL03PADEMD   5212718801
83021415:39:40.5215:41:13.24ADUGLL07PADDJR   5276960651
83021415:39:46.6215:41:24.84ADUGLL06PADJEB1  9987092501
83021415:40:16.9515:41:39.26ADUGLL03PADEMD   5212718801
83021415:39:46.6215:41:48.14ADUGLL06PADJEB1  9987092501
83021415:41:40.9315:42:30.63ADUGLL03PADEMD   1457063151
83021415:41:40.9315:42:48.41ADUGLL03PADEMD   1457063151
83021415:42:07.2215:43:23.46ADUGLL07PADDJR   3916409651
83021415:41:49.6415:43:25.79ADUGLL06PADJEB1  1085278721
83021415:42:07.2215:43:32.62ADUGLL07PADDJR   3916409651
83021415:41:49.6415:43:37.25ADUGLL06PADJEB1  1085278721
83021415:42:07.2215:44:31.64ADUGLL07PADDJR   3916409651
83021415:44:01.9415:44:44.65ADUGLL03PADEMD   2965454721
83021415:44:01.9415:44:56.54ADUGLL03PADEMD   2965454721
83021415:44:53.8115:46:07.35ADUGLL06PADJEB1  9987092511
83021415:44:33.0715:46:09.93ADUGLL07PADDJR   5269112791
83021415:44:53.8115:46:21.88ADUGLL06PADJEB1  9987092511
83021415:44:33.0715:46:45.93ADUGLL07PADDJR   5269112791
83021415:46:46.9515:47:52.19ADUGLL07PADDJR   5271708691
83021415:46:48.1615:48:06.17ADUGLL06PADJEB1  9987092521
83021415:46:46.9515:48:16.79ADUGLL07PADDJR   5271708692
83021415:46:48.1615:48:28.80ADUGLL06PADJEB1  9987092521
83021415:46:46.9515:48:34.92ADUGLL07PADDJR   5271708692
83021415:48:37.0815:49:43.73ADUGLL07PADDJR   5267730331
83021415:48:37.0815:49:48.30ADUGLL07PADDJR   5267730331
83021415:48:32.5315:50:17.17ADUGLL06PADJEB1  99870925310000010000000000000000000000000000000
83021415:48:32.5315:50:31.47ADUGLL06PADJEB1  9987092531
83021415:48:32.5315:50:49.70ADUGLL06PADJEB1  9987092531
83021415:48:37.0815:50:59.17ADUGLL07PADDJR   5267730331
83021415:50:54.8315:51:51.22ADUGLL06PADJEB1  5696340761
83021415:51:01.0115:51:51.30ADUGLL07PADDJR   5267506161
83021415:51:01.0115:52:26.71ADUGLL07PADDJR   5267506161
83021415:52:29.2915:53:12.34ADUGLL07PADDJR   5266145121
83021415:52:24.4915:53:28.19ADUGLL06PADJEB1  5268317491
83021415:52:29.2915:53:36.53ADUGLL07PADDJR   5266145121
83021415:52:24.4915:54:12.14ADUGLL06PADJEB1  5268317491
83021415:54:13.1915:55:49.25ADUGLL06PADJEB1  9987092541
83021415:54:13.1915:56:16.55ADUGLL06PADJEB1  9987092541
83021415:54:13.1915:56:23.57ADUGLL06PADJEB1  9987092541
83021415:56:24.9915:58:20.68ADUGLL06PADJEB1  99870925510000000000000000000000000000000000001
83021415:58:27.4815:59:50.69ADUGLL06PADJEB1  9987092561
83021415:58:27.4816:00:04.61ADUGLL06PADJEB1  9987092561
83021416:00:06.1316:01:00.64ADUGLL06PADJEB1  9987092571
```

RAW DATA

DATE	ASUID	TERMID	OPRID	COUNT	SESTIME	BEGTIME
8445	44683039	7	DJR	0	84.94	48399.9
8445	51646877	3	EMD	0	71.28	56344.4
8445	76583850	3	EMD	0	62.06	56280.0
8445	108527872	6	JEB1	0	107.61	56509.6
8445	118648526	5	HEM	0	84.20	48163.1
8445	124407231	7	JXF	0	149.89	52302.2
8445	130289703	6	CLS	0	71.15	50876.8
8445	135609932	6	MKT	0	71.15	48392.2
8445	145606721	6	MKT	0	135.54	47743.5
8445	145708315	3	EMD	0	67.48	56500.9
8445	146585351	7	KMR	0	81.84	49762.8
8445	158607818	7	JXF	0	95.09	52594.4
8445	180546506	6	MKT	0	135.00	47343.0
8445	221482586	6	CLS	0	108.82	51412.9
8445	228272087	7	KMR	0	141.61	49618.9
8445	264088932	7	JXF	0	98.46	51967.4
8445	296545472	3	EMD	0	54.60	56641.9
8445	300524132	3	EMD	1	77.40	55922.4
8445	321607972	7	JXF	0	122.03	51677.6
8445	335547770	6	MKT	0	109.72	48280.9
8445	346681943	7	DJR	0	87.42	56077.1
8445	352584465	5	HEM	0	94.48	48809.8
8445	357601132	5	HEM	0	124.94	48682.5
8445	362665113	3	EMD	0	84.40	56001.3
8445	378706848	5	HEM	0	103.40	48252.2
8445	382729021	6	MKT	0	135.05	47479.1
8445	389781769	3	EMD	1	86.20	51053.9
8445	391640965	7	DJR	0	144.42	56527.2
8445	454319034	7	JXF	0	71.20	52520.5
8445	474880804	7	DJR	1	120.94	47477.9
8445	497561711	5	HEM	0	133.92	49041.4
8445	500781234	7	DJR	0	93.77	56285.5
8445	501885643	5	HEM	0	217.29	49323.9
8445	513829765	7	DJR	0	124.97	47604.8
8445	519026469	5	HEM	0	105.28	48572.0
8445	521271880	3	EMD	0	82.31	56416.9
8445	521884460	7	JXF	0	108.51	51802.7
8445	522298203	7	JXF	0	127.37	52172.5
8445	522646936	6	CLS	0	93.08	51316.9
8445	522784246	6	CLS	0	99.86	51533.3
8445	523213623	6	MKT	0	107.14	47233.9
8445	523278558	3	EMD	0	69.38	55850.2

PREPARED DATA

Appendix H
Responses to Evaluation Survey

PHASE 1: ALL MONOCHROME

TERMGP	Monochrome					Monochrome			
Operator / Question	1	3	6	8	9	2	4	5	7
1	3	4	3	3	3	2	4	3	4
2	2	4	3	2	3	2	4	4	4
3	4	5	4	5	4	5	5	4	5
4	2	4	2	5	4	5	5	5	5
5	3	4	3	5	4	5	4	4	3
6	5	5	4	4	5	5	4	5	3
7	5	3	4	2	4	2	3	2	3
8	4	2	3	2	3	3	1	2	2
9	*	*	*	*	*	*	*	*	*
10	*	*	*	*	*	*	*	*	*
11	2	2	2	2	2	2	2	3	1

* Withheld to insure confidentiality

PHASE 2: COLOR VERSUS MONOCHROME

TERMGP	Monochrome					Color			
Operator / Question	1	3	6	8	9	2	4	5	7
1	4	4	3	3	3	3	4	3	4
2	2	4	2	3	3	1	3	2	1
3	4	4	3	5	4	5	4	4	3
4	2	4	2	5	4	4	4	4	4
5	3	4	4	5	4	5	4	3	3
6	5	5	4	4	5	5	5	5	2
7	4	3	4	2	3	5	4	3	5
8	3	2	3	1	3	2	2	1	3
9	*	*	*	*	*	*	*	*	*
10	*	*	*	*	*	*	*	*	*
11	3	2	2	2	3	2	3	2	4

* Withheld to insure confidentiality

PHASE 3: COLOR VERSUS MONOCHROME

TERMGP	Color					Monochrome			
Operator / Question	1	3	6	8	9	2	4	5	7
1	3	4	3	3	4	3	3	3	4
2	2	4	3	3	4	4	$	3	4
3	5	4	3	4	4	5	5	4	4
4	3	3	2	4	3	4	5	4	4
5	3	4	4	4	4	4	4	4	4
6	5	5	4	4	5	5	4	5	2
7	4	4	3	3	4	3	2	3	2
8	3	3	3	2	3	3	1	2	2
9	*	*	*	*	*	*	*	*	*
10	*	*	*	*	*	*	*	*	*
11	3	3	3	3	4	2	2	3	2

* Withheld to insure confidentiality